

# GENERALIZED FOURIER COEFFICIENTS OF MULTIPLICATIVE FUNCTIONS

LILIAN MATTHIESEN

ABSTRACT. We prove for a fairly general class of not necessarily bounded multiplicative functions that their elements give rise to functions that are orthogonal to polynomial nilsequences when applying a ‘ $W$ -trick’. The resulting functions therefore have small uniformity norms of all orders by the Green–Tao–Ziegler inverse theorem.

## CONTENTS

1. Introduction	1
2. Brief outline of some ideas	9
3. Digression on the Dirichlet decomposition of non-negative $f$	10
4. $W$ -trick	14
5. The non-correlation result	15
6. Linear subsequences of equidistributed nilsequences	23
7. Equidistribution of product nilsequences	29
7.1. Restriction of Proposition 7.1 to pairs of primes	33
8. Proof of Proposition 5.3	34
8.1. Reduction by hyperbola method	34
8.2. Contribution from exceptional $D$	36
8.3. Montgomery–Vaughan approach	38
8.4. Short sums	42
8.5. Large primes	43
8.6. Small primes	48
Appendix A. Explicit bounds on the correlation of $\Lambda$ with nilsequences	50
References	56

## 1. INTRODUCTION

Let  $f : \mathbb{N} \rightarrow \mathbb{R}$  be a multiplicative arithmetic function. Daboussi showed (see Daboussi and Delange [4]) that if  $|f|$  is bounded by 1, then

$$\frac{1}{x} \sum_{n \leq x} f(n) e^{2\pi i \alpha n} = o(x) \quad (1.1)$$

---

While working on this paper the author was supported by ERC grant 208091 and by a postdoctoral fellowship of the Fondation Sciences Mathématiques de Paris.

holds for every irrational  $\alpha$ . A detailed proof of the following slightly strengthened version may be found in Daboussi and Delange [5]: Suppose  $f$  satisfies

$$\sum_{n \leq x} |f(n)|^2 = O(x), \quad (1.2)$$

then (1.1) holds for every irrational  $\alpha$ . Montgomery and Vaughan [22] give explicit error terms for the decay in (1.1) for multiplicative functions that satisfy, in addition to (1.2), a uniform bound at all primes, in the sense that  $|f(p)| \leq H$  holds for some constant  $H \geq 1$  and all primes  $p$ .

In this paper we will study the closely related question of bounding correlations of multiplicative functions with polynomial nilsequences in place of the exponential function  $n \mapsto e^{2\pi i \alpha n}$ . A chief concern in this work is to include unbounded multiplicative functions in the analysis. To this end we shall significantly weaken the moment condition (1.2) by decomposing  $f$  into a suitable Dirichlet convolution  $f = f_1 * \cdots * f_t$  and analysing the correlations of the individual factors with exponentials, or rather nilsequences. The benefit of such a decomposition is that we merely require control on the second moments of the individual factors of the Dirichlet convolution and not of  $f$  itself. This essentially allows us to replace (1.2) by the condition that there exists  $\theta_f \in (0, 1]$  such that

$$\sqrt{\frac{1}{x} \sum_{n \leq x} |f_i(n)|^2} \ll (\log x)^{1-\theta_f} \frac{1}{x} \sum_{n \leq x} |f_i(n)| \quad (1.3)$$

for all  $i \in \{1, \dots, t\}$ . Before we continue, we provide an example of a function that satisfies (1.3), but neither (1.2) nor

$$\sum_{n \leq x} |f(n)|^2 \ll \sum_{n \leq x} |f(n)|. \quad (1.4)$$

**Example 1.1.** Let  $t > 1$  be an integer and consider the general divisor function  $d_t(n) = \mathbf{1} * \cdots * \mathbf{1}(n)$  which is defined to be the  $t$ -fold convolution of  $\mathbf{1}$  with itself. This is a function which satisfies (1.3), but neither (1.2) nor (1.4). Indeed, we have

$$\frac{1}{x} \sum_{n \leq x} d_t(n) \asymp_t (\log x)^{t-1},$$

while the second moment satisfies

$$\frac{1}{x} \sum_{n \leq x} d_t^2(n) \asymp_t (\log x)^{t^2-1}.$$

Thus, the second moment is not controlled by the first. On the other hand, choosing  $f_i = \mathbf{1}$  for each  $1 \leq i \leq t$ , it is clear that (1.3) holds with  $\theta_f = 1$ .

To define the class of functions we shall work with, we introduce the following notation. For integers  $q, r \in \mathbb{N}$  let

$$S_f(x; q, r) = \frac{q}{x} \sum_{\substack{n \leq x \\ x \equiv r \pmod{q}}} f(x)$$

denote the average value of  $f$  on the progression  $x \equiv r \pmod{q}$ . Let  $w : \mathbb{N} \rightarrow \mathbb{R}$  be an increasing function such that  $\frac{\log \log x}{\log \log \log x} < w(x) \leq \log \log x$  for all  $x > e^e$  and set

$$W(x) = \prod_{p \leq w(x)} p.$$

**Definition 1.2.** Given positive integers  $E$  and  $H \geq 1$ , we let  $\mathcal{M}_H(E)$  denote the class of multiplicative arithmetic functions  $f : \mathbb{N} \rightarrow \mathbb{R}$  with the following property: There is a decomposition  $f = f_1 * \cdots * f_t$  for some  $t \leq H$  such that each  $f_i$  satisfies:

- (1)  $|f_i(p^k)| \leq H^k$  for all prime powers  $p^k$ .
- (2) There is  $\theta_f \in (0, 1]$  such that for all  $x > 1$ , for all positive integers  $q \leq (\log x)^E$  and for all  $A \in (\mathbb{Z}/Wq\mathbb{Z})^*$  with  $S_f(x; Wq, A) > 0$  we have

$$\sum_{i=1}^t \sum_{\substack{d_1, \dots, d_i, \dots, d_t \\ (D_i, W)=1 \\ D_i \leq x^{1-1/t}}} \left( \prod_{j \neq i} \frac{|f_j(d_j)|}{d_j} \right) \sqrt{\frac{D_i W q}{x} \sum_{\substack{n \leq x/D_i \\ n D_i \equiv A \pmod{Wq}}} |f_i(n)|^2} \\ \ll (\log x)^{1-\theta_f} S_{|f|}(x; Wq, A),$$

as  $x \rightarrow \infty$ , where  $D_i = \prod_{j \neq i} d_j$  and  $D_1 = 1$  if  $t = 1$ , and  $W = W(x)$ .

- (3)  $|f_i|$  has a stable mean value in the following sense: whenever  $x \ll x'^{C_1} \ll x^{C_2}$  for positive constants  $C_1$  and  $C_2$ , then

$$S_{|f_i|}(x; W(x)q, A) \asymp_{C_1, C_2} S_{|f_i|}(x'; W(x)q, A)$$

for all positive integers  $q \leq (\log x)^E$  and for all  $A \in (\mathbb{Z}/W(x)q\mathbb{Z})^*$  such that  $S_f(x; W(x)q, A) > 0$ .

- (4) For all  $x > 1$ , for all positive integers  $q \leq (\log x)^E$  and all residues  $A \in (\mathbb{Z}/W(x)q\mathbb{Z})^*$  such that  $S_f(x; W(x)q, A) > 0$ , we have

$$\frac{1}{\log x} \sum_{\substack{j \in \mathbb{N}: \\ \exp((\log \log x)^2) < 2^j \leq x}} S_{|f_i|}(2^j; W(x)q, A) \ll S_{|f_i|}(x; W(x)q, A).$$

In other words, if  $P$  denotes the arithmetic progression  $\{n \equiv A \pmod{W(x)q}\}$ , then the average of  $|f_i|$ -mean values on all sufficiently long subprogressions of the form  $P \cap [1, 2^j] \subseteq P \cap [1, x]$  does, up to a constant factor, not exceed that of the mean value with respect to  $P \cap [1, x]$ .

In view of this rather long definition we discuss in the sequel three simple examples of functions that belong to  $\mathcal{M}_H(E)$  and provide a few general remarks on the individual conditions (1)-(4) before presenting in more detail the actual aim of this paper.

The examples we consider are the general divisor function  $d_H$  (that is, the  $H$ -fold convolution of  $\mathbf{1}$  with itself), the characteristic function  $\mathbf{1}_S$  of the set of sums of two squares, and the representation function  $r$  of sums of two squares. Given any positive constant  $E$ , we show below, that  $d_H \in \mathcal{M}_H(E)$ ,  $\mathbf{1}_S \in \mathcal{M}_1(E)$ , and  $r \in \mathcal{M}_2(E)$ .

**Example 1.3.** Let  $f = d_H$ , and set  $f_i = \mathbf{1}$  for  $1 \leq i \leq H$ . We check that  $f \in \mathcal{M}_H(E)$  for any  $E = O(1)$ . Indeed, (1) holds trivially. Since  $S_{f_i}(x; q, r) \asymp 1$ , conditions (3) and (4) are also easily checked. Condition (2) can be checked in a similar way as the special case considered in Example 1.1.

**Example 1.4.** Let  $S = \{x^2 + y^2 \mid x, y \in \mathbb{Z}\}$  denote the set of integers that are representable as a sum of two squares and let  $f = \mathbf{1}_S$  be its characteristic function. Then  $f \in \mathcal{M}_1(E)$  for all  $E = O(1)$ . Indeed,  $f$  is multiplicative and clearly satisfies (1). Prachar [24] showed that for  $a \equiv 1 \pmod{\gcd(4, q)}$  we have

$$\sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} \mathbf{1}_S(n) \sim B_q x (\log x)^{-1/2}, \quad (1.5)$$

where

$$B_q = B q^{-1} \frac{(4, q)}{(2, q)} \prod_{\substack{p \equiv 3 \pmod{4} \\ p \mid q}} (1 + p^{-1}) \quad \text{and} \quad B = \frac{1}{\sqrt{2}} \prod_{p \equiv 3 \pmod{4}} (1 - p^{-2})^{-1/2}.$$

Thus, (3) holds since  $(\log x)^{-1/2} \asymp (\log x')^{-1/2}$ . Condition (4) is satisfied since

$$\begin{aligned} \sum_{j=\lfloor (\log \log x)^2 / \log 2 \rfloor}^{\log x / \log 2} (\log 2^j)^{-1/2} &= \sum_{j=\lfloor (\log \log x)^2 / \log 2 \rfloor}^{\log x / \log 2} (j \log 2)^{-1/2} \\ &\ll (\log x)^{1-1/2} - (\log \log x) \ll (\log x)^{1-1/2}. \end{aligned}$$

To check (2), suppose that  $A \in (\mathbb{Z}/Wq\mathbb{Z})^*$  is representable as a sum of two squares and observe that this implies

$$1 \ll \phi(Wq) \frac{D_i}{x} \sum_{\substack{p \leq x/D_i \\ p \equiv A \pmod{Wq}}} \mathbf{1}_S(p) \log p \ll (\log x) \frac{WqD_i}{x} \sum_{\substack{n \leq x/D_i \\ n \equiv A \pmod{Wq}}} \mathbf{1}_S(n).$$

Thus,

$$\begin{aligned} \frac{WqD_i}{x} \sum_{\substack{n \leq x/D_i \\ n \equiv A \pmod{Wq}}} \mathbf{1}_S^2(n) &= \frac{WqD_i}{x} \sum_{\substack{n \leq x/D_i \\ n \equiv A \pmod{Wq}}} \mathbf{1}_S(n) \\ &\ll \log x \left( \frac{WqD_i}{x} \sum_{\substack{n \leq x/D_i \\ n \equiv A \pmod{Wq}}} \mathbf{1}_S(n) \right)^2. \end{aligned}$$

Inserting this into the left hand side of (2) with  $D_1 = 1$ , we see that (2) is satisfied with  $\theta_f = 1/2$ .

**Example 1.5.** As a final example, we consider the function

$$f(n) = r(n) := \frac{1}{4} \# \left\{ (x, y) : x^2 + y^2 = n \right\}$$

and show that  $f \in \mathcal{M}_2(E)$  for all  $E = O(1)$ . We decompose  $r$  as  $r = \mathbf{1}'_S * \mathbf{1}''_S$ , where  $\mathbf{1}'_S(n) = \mu^2(n)\mathbf{1}_S(n)\mathbf{1}_{2|n}$  for the function  $\mathbf{1}_S$  from the previous example, and  $\mathbf{1}''_S$  is chosen in such a way that the convolution identity holds. In particular,  $\mathbf{1}''_S(n)\mu^2(n) = \mathbf{1}_S(n)\mu^2(n)$ . It is not difficult to check that both functions  $\mathbf{1}'_S$  and  $\mathbf{1}''_S$  satisfy condition (1). In both cases,  $f_i = \mathbf{1}'_S$  and  $f_i = \mathbf{1}''_S$ , we have

$$\frac{\phi(Wq)D_i}{x} \sum_{\substack{p \leq x/D_i \\ p \equiv A' \pmod{Wq}}} f_i(p) \log p \gg 1$$

whenever  $A' \in (\mathbb{Z}/Wq\mathbb{Z})^*$  is a sum of two squares. Furthermore, either of the conditions  $\mathbf{1}'_S(n) > 0$  and  $\mathbf{1}''_S(n) > 0$  implies  $\mathbf{1}_S(n) > 0$ . Thus, when  $A \in (\mathbb{Z}/Wq\mathbb{Z})^*$  is a sum of two squares and when  $\mathbf{1}'_S(D_1) > 0$  or  $\mathbf{1}''_S(D_1) > 0$  for some  $D_1$  with  $D_1\overline{D_1} \equiv 1 \pmod{Wq}$ , then  $\overline{D_1}A \in (\mathbb{Z}/Wq\mathbb{Z})^*$  is a sum of two squares. Thus, let  $A' \in (\mathbb{Z}/Wq\mathbb{Z})^*$  be a sum of two squares. Then

$$\begin{aligned} \frac{WqD_i}{x} \sum_{\substack{n \leq x/D_i \\ n \equiv A' \pmod{Wq}}} f_i(n)^2 &= \frac{WqD_i}{x} \sum_{\substack{n \leq x/D_i \\ n \equiv A' \pmod{Wq}}} \mathbf{1}_S(n) \\ &\ll \log x \left( \frac{WqD_i}{x} \sum_{\substack{n \leq x/D_i \\ n \equiv A' \pmod{Wq}}} \tilde{f}_i(n) \right)^2, \end{aligned}$$

where  $\tilde{f}_i$  is multiplicatively defined via  $\tilde{f}_i(p) = f_i(p)$  at primes and  $\tilde{f}_i(p^k) = f_i(p^k)^2$  when  $k \geq 2$ . Condition (2) can now be deduced with the help of Shiu's lemma (cf. Lemma 3.3). We refer to the proof of Lemma 3.1 for more details. Properties (3) and (4) follow in a similar way as in the previous example. The trick is that

$$S_{\mathbf{1}_S}(x; Wq, A) \asymp S_h(x; Wq, A)$$

for  $h(n) = \mu^2(n)\mathbf{1}_S(n)$ , which allows us to obtain lower bounds on  $S_{f_i}(x; Wq, A)$  that match the upper bounds provided by Shiu's lemma.

**Remarks on the individual conditions in Definition 1.2.** (i) If  $f$  is a function that enjoys good second moment estimates, then the parameter  $t$  in Definition 1.2 can be chosen to equal 1 in order to ensure that condition (2) holds. It is not difficult to show that any bounded multiplicative function  $f$  satisfies the condition (2) for  $t = 1$  if it satisfies density conditions of the form

$$\sum_{\substack{p \leq x \\ p \equiv A \pmod{Wq}}} |f(p)| \log p \gg \frac{x}{\phi(Wq)}.$$

More generally, we shall show in Lemma 3.1 that any non-negative  $f : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$  that satisfies  $f(p^k) \leq H^k$  at prime powers admits a Dirichlet decomposition  $f = f_1 * \cdots * f_H$  with properties (1) and (2) if it satisfies density conditions on the set of primes like the one

above and if it has Shiu-type average values in arithmetic progressions. The decomposition we work with in Lemma 3.1 is given by

$$f_i(p^k) = \begin{cases} f(p)/H & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}$$

for  $1 \leq i < H$ , and  $f_H$  is defined in such a way that  $f = f_1 * \cdots * f_H$  holds.

(ii) Observe that condition (3) is not a strong restriction, since the bound  $|f_i(p^k)| \leq (CH)^k$  implies (cf. Lemma 3.3) that

$$S_{|f_i|}(x; Wq, r) \ll (\log x)^{O(H)}$$

for any positive integer  $q \leq x^{1/2}$  and  $r \in (\mathbb{Z}/Wq\mathbb{Z})^*$ .

(iii) Setting  $R(t) = S_{|f_i|}(2^t; Wq, r)$ , then condition (4) may be written as

$$\int_{(\log y)^2}^y R(t) dt \ll yR(y).$$

If  $R$  was differentiable, this would translate to  $R(t) \ll R(t) + tR'(t)$ . That is, there is  $c > 0$  such that  $-R'(t) < (1 - c)(R(t)/t)$ , or, equivalently,  $\frac{R'(t)}{R(t)} > (-1 + c)\frac{1}{t}$ . Integrating yields that for each solution there are constants  $C_1 > 0$  and  $C_2$  such that  $R(t) > C_1 t^{-1+c} + C_2$ , which essentially means that  $|f_i|$  is strictly denser than the set of primes.

In view of the final observation from above, we may without imposing further restrictions add the following assumption to the list from Definition 1.2. This will lead to a slight simplification of the error terms we obtain in Section 5 (cf. Section 8.2).

**Definition 1.6.** Let  $\mathcal{M}_H^*(E)$  denote the class of multiplicative functions  $f \in \mathcal{M}_H(E)$  for which the decomposition  $f = f_1 * \cdots * f_t$  from Definition 1.2 has the additional property that there is a positive constant  $\alpha_f > 0$  such that for all  $x > 1$  and all positive integers  $q \leq (\log x)^E$  the following estimate holds:

$$\max_{1 \leq i \leq t} \max_{A \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|f_i|}(x; \widetilde{W}q, A) \gg (\log x)^{-1+\alpha_f}. \quad (1.6)$$

**Aim and motivation.** As mentioned before, the purpose of this paper is to study correlations of multiplicative functions, more specifically of functions from  $\mathcal{M}_H^*(E)$ , with polynomial nilsequences. In general, such correlations can only be shown to be small if either the nilsequence is highly equidistributed or else if the multiplicative function is equidistributed in progressions with short common difference. We will consider both cases, the former in Proposition 5.3 and the latter in Theorem 5.1. Taking account of the above mentioned restriction, the latter result applies to the subset  $\mathcal{F}_H(E) \subset \mathcal{M}_H^*(E)$  of functions that satisfy the *major arc condition* that we will introduce as (4.1) in Section 4. In short, there is for every  $f \in \mathcal{F}_H(E)$  a product  $\widetilde{W} = \widetilde{W}(x)$  of small prime powers such that  $f$  has a constant average value on certain subprogressions of the progression  $\{n \equiv A \pmod{\widetilde{W}}\}$

for any fixed residue  $A \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*$ . Instead of the bound (1.1) on Fourier coefficients of  $f$ , we aim to show that every  $f \in \mathcal{F}_H(E)$  satisfies

$$\begin{aligned} & \frac{\widetilde{W}}{x} \sum_{n \leq x/\widetilde{W}} \left( f(\widetilde{W}n + A) - S_f(x; \widetilde{W}, A) \right) F(g(n)\Gamma) \\ &= o_{G/\Gamma} \left( \frac{1}{\log x} \exp \left( \sum_{w(x) < p \leq x} \frac{|f_1(p)| + \cdots + |f_t(p)|}{p} \right) \right) \end{aligned} \quad (1.7)$$

for all 1-bounded polynomial nilsequences  $F(g(n)\Gamma)$  of bounded degree and bounded Lipschitz constant that are defined with respect to a nilmanifold  $G/\Gamma$  of bounded step and bounded dimension. The precise statement will be given in Section 5.

The interest in estimates of the form (1.7) lies in the fact that the Green–Tao–Ziegler inverse theorem [15] allows one to deduce that  $f(\widetilde{W}n + A) - S_f(x; \widetilde{W}, A)$  has small  $U^k$ -norms of all orders, where ‘small’ may depend on  $k$ . Employing the nilpotent Hardy–Littlewood method of Green and Tao [12], this in turn allows one to deduce asymptotic formulae for expressions of the form

$$\sum_{\mathbf{x} \in K \cap \mathbb{Z}^s} f(\varphi_1(\mathbf{x}) + a_1) \cdots f(\varphi_r(\mathbf{x}) + a_r), \quad (1.8)$$

where  $a_1, \dots, a_r \in \mathbb{Z}$ , where  $\varphi_1, \dots, \varphi_r : \mathbb{Z}^s \rightarrow \mathbb{Z}$  are pairwise non-proportional linear forms, and where  $K \subset \mathbb{R}^s$  is convex, *provided* that  $f$  has a sufficiently pseudorandom majorant. Such majorants were constructed for a general class of positive multiplicative functions with divisor function like growth conditions in [3, §7].

We note aside that in the special case where the nilsequence is replaced by  $e^{2\pi i n \alpha}$  for irrational  $\alpha$  the approach of this paper yields for  $f \in \mathcal{M}_H(E)$  with  $E = O(1)$  that

$$\frac{1}{x} \sum_{n \leq x} f(n) e^{2\pi i n \alpha} = o \left( \frac{1}{x} \sum_{n \leq x} |f_1| * \cdots * |f_t|(n) \right).$$

(Here, no  $W$ -trick is required, since the sequence  $n \mapsto e^{2\pi i n \alpha}$  is equidistributed and we have  $\int_{\mathbb{R}/\mathbb{Z}} e^{2\pi i t} dt = 0$ .) Restricting to non-negative functions  $f$  for which  $\sum_{n \leq x} f(n) \asymp \sum_{n \leq x} |f_1| * \cdots * |f_t|(n)$ , this provides a new family of examples of multiplicative functions  $f$  that satisfy the condition of a question due to Dupain, Hall and Tenenbaum [6, p.410]. We refer to Bachman [1] for further results and a more detailed discussion of the Dupain–Hall–Tenenbaum question.

**Strategy and related work.** Our overall strategy is to decompose the given multiplicative function via Dirichlet decomposition in such a way that we can employ the Montgomery–Vaughan approach to the individual factors. This approach reduces matters to bounding correlations of sequences defined in terms of primes. We apply Green and Tao’s result [12, Prop. 10.2] on the correlation of the ‘ $W$ -tricked von Mangoldt function’ with nilsequences to bound one type of correlation that appears. Carrying out the Montgomery–Vaughan approach in the nilsequences setting furthermore necessitates the

understanding of equidistribution properties of certain families of product nilsequences that arise after an application of the Cauchy–Schwarz inequality. These products are studied in Section 7 refining techniques introduced in [14]. More precisely, we show that most of these products are equidistributed when the original sequence from which the products are derived is equidistributed. The latter can be achieved by the Green–Tao factorisation theorem for nilsequences from [13].

The question studied in this paper is in spirit related to that of Bourgain–Sarnak–Ziegler [2], who use an orthogonality criterion that can be proved employing ideas that go back to Daboussi–Delange [4] (cf. Harper [16] and Tao [26]). Invoking the orthogonality criterion in the form it is presented in Kátai [19], recent and very substantial work of Frantzikinakis and Host [9] shows that every bounded multiplicative function can be decomposed into the sum of a Gowers-uniform function, a structured part and an error term. Here, the error term is small in the sense that the integral of the error term over the space of all 1-bounded multiplicative functions is small. While this result provides no information on the quality of the error term of individual functions, it allows one to study simultaneously all bounded multiplicative functions.

The point of view taken in the present work is a different one: we have applications to explicit multiplicative functions in mind that appear naturally in number theoretic contexts. Such functions will often have a mean value  $\frac{1}{x} \sum_{n \leq x} f(n)$  that is given by a reasonably nice function in  $x$ . We assume here that for our explicit function  $f$  the conditions from Definition 1.2 can be verified. In order to deduce asymptotic formulae for expressions of the form (1.8), it is indispensable to understand the error terms in the non-correlation result well enough in order to see that they improve on the trivial bound given by the average value of  $|f|$ . The non-correlation result (Theorem 5.1) we establish comes with explicit error terms which preserve at least some information on the average value of the multiplicative function. An important feature of this work is that it applies to a large class of unbounded functions.

The results of this paper have several natural applications involving for instance the functions considered in Examples 1.3–1.5 but also more interesting functions like the absolute value of eigenvalues of holomorphic cusp forms. We will discuss these applications in forthcoming work.

**Notation.** The following, perhaps unusual, piece of notation will be used throughout the paper: Suppose  $\delta \in (0, 1)$ , then we write  $x = \delta^{-O(1)}$  instead of  $x = (1/\delta)^{O(1)}$  to indicate that there is a constant  $0 \leq C \ll 1$  such that  $x = (1/\delta)^C$ .

**Convention.** If the statement of a result contains Vinogradov or  $O$ -notation in the assumptions, then the implied constants in the conclusion may depend on all implied constants from the assumptions.

**Acknowledgements.** I would like to thank Hedi Daboussi, Régis de la Bretèche, Nikos Frantzikinakis, Andrew Granville, Ben Green, Dimitris Koukoulopoulos and Terry Tao for very helpful discussions and Nikos Frantzikinakis for valuable comments on a previous version of this paper.



## 2. BRIEF OUTLINE OF SOME IDEAS

In this section we give a very rough outline of the ideas behind the application of the Montgomery–Vaughan approach. This section is only intended as guidance. In particular, we make a number of oversimplifications, so that statements will not be quite true as stated here. The main idea of Montgomery–Vaughan [22] is to introduce log factor into the Fourier coefficient that we wish to analyse. Let  $f : \mathbb{N} \rightarrow \mathbb{R}$  be a multiplicative function that satisfies  $|f(p)| \leq H$  for some constant  $H \geq 1$  and all primes  $p$  and suppose (1.4) holds. Then we have

$$\sum_{n \leq N} f(n) e(n\alpha) \log(N/n) \leq \left( \sum_{n \leq N} (\log N/n)^2 \right)^{1/2} \left( \sum_{n \leq N} |f(n)|^2 \right)^{1/2} \ll N^{1/2} \left( \sum_{n \leq N} |f(n)|^2 \right)^{1/2},$$

and thus

$$\log N \left( \frac{1}{N} \sum_{n \leq N} f(n) e(n\alpha) \right) \ll \left( \frac{1}{N} \sum_{n \leq N} |f(n)|^2 \right)^{1/2} + \left| \frac{1}{N} \sum_{n \leq N} f(n) e(n\alpha) \log n \right|.$$

The first term in the bound is handled by the assumptions on  $f$ , that is, by assuming that (1.4) holds. To bound the second term, one invokes the identity  $\log n = \sum_{d|n} \Lambda(d)$ , which reduces the task to bounding the expression

$$\sum_{nm \leq N} f(nm) \Lambda(m) e(nm\alpha).$$

This in turn may be reduced to the task of bounding

$$\sum_{np \leq N} f(n) f(p) \Lambda(p) e(pn\alpha),$$

where  $p$  runs over primes. Applying the Cauchy-Schwarz inequality and smoothing, it furthermore suffices to estimate expressions of the form

$$\sum_{p, p'} f(p) f(p') \log(p) \log(p') \sum_n w(n) e((p - p')n\alpha),$$

where  $p$  and  $p'$  run over primes and where  $w$  is a smooth weight function. One employs a standard sieve estimate to bound  $\#\{(p, p') : p - p' = h\}$  for fixed  $h$ . Standard exponential sum estimates and a delicate decomposition of the summation ranges for  $n, p, p'$  yield an explicit bound on  $\frac{1}{N} \sum_{n \leq N} f(n) e(n\alpha)$ .

It will be our aim to employ the above approach to correlations of the form

$$\frac{1}{N} \sum_{n \leq N} \left( f(n) - \frac{1}{N} \sum_{m \leq N} f(m) \right) F(g(n)\Gamma)$$

for multiplicative  $f$ . One problem we face is that the above approach makes substantial use of the strong equidistribution properties of the exponential functions  $e((p - p')n\alpha)$  for distinct primes  $p, p'$ . A general nilsequence  $F(g(n)\Gamma)$  on the other hand is not even equidistributed. This problem is resolved by an application of the factorisation theorem for polynomial sequences from Green–Tao [13], which allows us to assume that  $(g(n)\Gamma)_{n \leq N}$

is equidistributed in  $G/\Gamma$  if  $f$  is equidistributed in progressions to small moduli. The latter will be arranged for by employing a  $W$ -trick. As above, we then consider the following expression which we split into the case of large, resp. small primes with respect to a suitable cut-off parameter  $X$ :

$$\begin{aligned} \frac{1}{N} \sum_{mp \leq N} f(m)f(p)\Lambda(p)F(g(mp)\Gamma) &= \frac{1}{N} \sum_{m \leq X} \sum_{p \leq N/m} f(m)f(p)\Lambda(p)F(g(mp)\Gamma) \\ &+ \frac{1}{N} \sum_{m > X} \sum_{p \leq N/m} f(m)f(p)\Lambda(p)F(g(mp)\Gamma). \end{aligned}$$

Applying Cauchy-Schwarz to both terms shows that it suffices to understand correlations of the form

$$\sum_{m, m'} f(m)f(m') \sum_p \Lambda(p)F(g(mp)\Gamma) \overline{F(g(m'p)\Gamma)}$$

and

$$\sum_{p, p'} f(p)f(p')\Lambda(p)\Lambda(p') \sum_m F(g(pm)\Gamma) \overline{F(g(p'm)\Gamma)}.$$

Choosing  $X$  suitably, only the first of these correlations matters. We shall bound this correlation by employing Green and Tao's result that the  $W$ -tricked von Mangoldt function is orthogonal to nilsequences. The necessary equidistribution properties of the sequences  $n \mapsto F(g(mn)\Gamma) \overline{F(g(m'n)\Gamma)}$  will be established in Sections 6 and 7.

We explain at the beginning of Section 8 how the Dirichlet decomposition  $f = f_1 * \cdots * f_t$  from Definition 1.2 can be employed to use this method for functions  $f$  that enjoy property (2) of Definition 1.2 in place of the moment condition (1.4).

### 3. DIGRESSION ON THE DIRICHLET DECOMPOSITION OF NON-NEGATIVE $f$

The Montgomery–Vaughan approach outlined in the previous section requires good control of the  $L^2$ -norm of the multiplicative function  $f$  in terms of its mean value. Let  $N$  and  $T$  be positive integer parameters that satisfy  $N^{1-o(1)} \ll T \ll N$ , and write  $W = W(N)$ . A function  $f : \mathbb{N} \rightarrow \mathbb{R}$  will exhibit a sufficient amount of control if it satisfies the bound

$$\sqrt{\frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} |f(n)|^2} \ll (\log N)^{1-\theta} \frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} |f(n)| \quad (3.1)$$

for all positive integers  $q \leq (\log N)^{O(1)}$ , for all  $A' \in (\mathbb{Z}/Wq\mathbb{Z})^*$ , and for some  $\theta \in (0, 1]$ . However, the condition (3.1) is rather strong and excludes many interesting examples of multiplicative functions such as the divisor function or the function that counts representations as sums of two squares. In order to extend the class of multiplicative functions that we can handle, we consider decompositions of  $f$  as the Dirichlet convolution  $f = f_1 * \cdots * f_t$  of a bounded number of multiplicative functions  $f_i : \mathbb{N} \rightarrow \mathbb{R}$  such that the  $L^2$ -norms of the  $f_i$  are controlled on average by the mean value of  $f$ . For given integers  $d_1, \dots, d_t$

and  $1 \leq i \leq t$ , we write  $D_i := \prod_{j \neq i} d_j$ . Then the following condition forms a suitable generalisation of (3.1):

$$\begin{aligned} \sum_{i=1}^t \sum_{\substack{d_1, \dots, \hat{d}_i, \dots, d_t \\ (D_i, W)=1 \\ D_i \leq T^{1-1/t}}} \left( \prod_{j \neq i} \frac{|f_j(d_j)|}{d_j} \right) \sqrt{\frac{D_i W q}{T} \sum_{\substack{n \leq T/D_i \\ n D_i \equiv A' \pmod{Wq}}} |f_i(n)|^2} \\ \ll (\log T)^{1-\theta_f} \frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} |f(n)| \end{aligned} \quad (3.2)$$

for some quantity  $\theta_f \in (0, 1]$ . Observe that the bound on  $D_i$  guarantees that every second moment that occurs on the left hand side has a long summation range. The special case  $t = 1$  corresponds to the bound (3.1).

Since the condition (3.2) might look rather restrictive, the purpose of this section is to give some evidence for the applicability of this approach. The following lemma provides sufficient conditions for the existence of Dirichlet decompositions  $f = f_1 * \dots * f_t$  with the property that (3.2) holds.

**Lemma 3.1.** *Let  $E$  and  $H$  be positive integers. Let  $f$  be a multiplicative function that satisfies  $|f(p^k)| \leq H^k$  at all prime powers  $p^k$ . Suppose that  $N$ ,  $T$  and  $W$  are as before, and suppose as well that for all positive integers  $q \leq (\log N)^E$  and for all  $A' \in (\mathbb{Z}/Wq\mathbb{Z})^*$  such that  $S_{|f|}(x; Wq, A') > 0$  the following two conditions are satisfied:*

$$\sum_{\substack{p \leq T \\ p \equiv A' \pmod{Wq}}} |f(p)| \log p \gg T/\phi(Wq) \quad (3.3)$$

and

$$S_{|f|}(x; Wq, A') \asymp \frac{1}{\phi(Wq)} \frac{1}{\log x} \exp \left( \sum_{\substack{p \leq x \\ p \nmid Wq}} \frac{|f(p)|}{p} \right). \quad (3.4)$$

Then  $f$  admits a Dirichlet decomposition  $f = f_1 * \dots * f_H$  of real-valued multiplicative functions  $f_i$  such that the condition (3.2) holds for  $t = H$  and  $\theta_f = \frac{1}{2}$ .

**Remark 3.2.** Note that (3.4) agrees with the bound predicted by Shiu's lemma. Assuming all the conditions of the lemma except for (3.4), Wirsing's result [27, Satz 1.1] implies that

$$\sum_{n \leq x} |f(n)| \asymp \frac{x}{\log x} \exp \left( \sum_{p \leq x} \frac{|f(p)|}{p} \right).$$

Thus, condition (3.4) is satisfied on a positive proportion of residues  $A' \in (\mathbb{Z}/W(N)q\mathbb{Z})^*$ .

Before we prove the lemma, we record a quick consequence of Shiu [25, Theorem 1] that we shall make use of.

**Lemma 3.3** (Shiu). *Let  $H$  be a positive integer and suppose  $f : \mathbb{N} \rightarrow \mathbb{R}$  is a non-negative multiplicative function satisfying  $f(p^k) \leq H^k$  at all prime powers  $p^k$ . Let  $W = W(x)$  be as before, let  $q > 0$  be an integer and let  $A' \in (\mathbb{Z}/Wq\mathbb{Z})^*$ . Then, as  $x \rightarrow \infty$ ,*

$$\sum_{\substack{0 < n \leq x \\ n \equiv A' \pmod{Wq}}} f(n) \ll \frac{x}{\phi(Wq) \log x} \exp \left( \sum_{\substack{w(x) < p \leq x \\ p \nmid q}} \frac{f(p)}{p} \right), \quad (3.5)$$

uniformly in  $A'$  and  $q$ , provided that  $q \leq x^{1/2}$ .

*Proof.* This lemma differs from [25, Theorem 1] in that it does not concern short intervals but at the same time it does not require  $f$  to satisfy  $f(n) \ll_\varepsilon n^\varepsilon$ . Shiu's result works with a summation range  $x - y < n \leq x$  where  $x^\beta < y \leq x$ ,  $\beta \in (0, \frac{1}{2})$ . Here, we are only interested in the case  $x = y$ . Thus, the parameter  $\beta$  can be regarded as fixed. As observed in [23], the proof of [25, Theorem 1] only requires the condition  $f(n) \ll_\varepsilon n^\varepsilon$  to hold for one fixed value of  $\varepsilon$  once  $\beta$  is fixed.

Note that any integer  $n \equiv A' \pmod{Wq}$  is free from prime divisors  $p < w(x)$ . Thus,  $f(n) \leq H^{\Omega(n)} \leq n^{\log H / \log w(x)}$ . Given any  $\varepsilon > 0$ , we deduce that  $n \equiv A' \pmod{Wq}$  implies  $f(n) \leq n^\varepsilon$  provided  $x$  is sufficiently large.  $\square$

*Proof of Lemma 3.1.* We define  $f_i$  for  $1 \leq i \leq H - 1$  multiplicatively by setting

$$f_i(p^k) = \begin{cases} f(p)/H & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}.$$

The remaining function  $f_H$  is defined in such a way that  $f = f_1 * \cdots * f_H$  holds true. Thus, we have  $f_H(p) = f(p)/H$  as well and the values at higher prime powers  $p^k$  are inductively defined via  $f(p^k) = f_1 * \cdots * f_H(p^k)$ . This yields

$$\begin{aligned} f_H(p^k) = f(p^k) - & \left( f_H(p^{k-1}) \frac{f(p)}{H} (H-1) + f_H(p^{k-2}) \left( \frac{f(p)}{H} \right)^2 \binom{H-1}{2} + \cdots \right. \\ & \left. + f_H(p^{k-H+1}) \left( \frac{f(p)}{H} \right)^{H-1} \binom{H-1}{H-1} \right). \end{aligned}$$

We claim that  $|f_H(p^k)| \ll (3H)^k$ . Indeed, for  $k = 1$  we have  $f(p)/H \leq 1$ . An induction hypothesis of the form  $|f_H(p^j)| \ll (3H)^j$  for all  $j < k$  leads to the conclusion that

$$|f_H(p^k)| \ll H^k + 3^{k-1} H^k \left( 1 + \frac{1}{2!} + \cdots + \frac{1}{(H-1)!} \right) \ll (3H)^k.$$

Note that the functions  $|f_1|, \dots, |f_{H-1}|$  take values in  $[0, 1]$ . Hence,

$$\frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} f_i^2(n) \leq \frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} |f_i(n)|. \quad (1 \leq i \leq H-1)$$

Furthermore, we have

$$\frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} |f_i(n)| \ll \frac{\log T}{\log w(N)} \left( \frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} |f_i(n)| \right)^2,$$

since

$$\frac{\log T}{T} \frac{Wq}{\log w(N)} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} |f_i(n)| \geq \frac{\phi(Wq)}{T} \sum_{\substack{p \leq T \\ p \equiv A' \pmod{Wq}}} |f_i(p)| \log p \gg_H 1,$$

by the assumptions on  $f$ .

In the case of  $f_H$ , we consider the auxiliary function  $\tilde{f}_H : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$ , defined multiplicatively by

$$\tilde{f}_H(p^k) = \begin{cases} |f_H(p^k)| & \text{if } p < w(N), \\ |f_H(p)| & \text{if } k = 1, p > w(N), \\ f_H(p^k)^2 & \text{if } k > 1, p > w(N). \end{cases}$$

Then, similarly as before,

$$\frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} f_H^2(n) \leq \frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} \tilde{f}_H(n) \ll \log T \left( \frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} \tilde{f}_H(n) \right)^2.$$

Thus, setting  $\tilde{f}_i = |f_i|$  for  $1 \leq i < H$ , we obtain an upper bound for the left hand side of (3.2) of the form

$$\begin{aligned} &\ll \frac{1}{(\log T)^{1/2}} \sum_{i=1}^H \sum_{\substack{d_1, \dots, \hat{d}_i, \dots, d_H \\ (D_i, W)=1}} \left( \prod_{j \neq i} \frac{\tilde{f}_j(d_j)}{d_j} \right) \frac{D_i Wq}{T} \sum_{\substack{n \leq T/D_i \\ n D_i \equiv A' \pmod{Wq}}} \tilde{f}_i(n) \\ &\ll_H \frac{Wq}{T(\log T)^{1/2}} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} \tilde{f}_1 * \dots * \tilde{f}_H(n). \end{aligned}$$

The functions  $\tilde{f}_1 * \dots * \tilde{f}_H$  and  $|f|$  agree at square-free integers  $n$  and the remaining difficulty is to bound the contribution of values at integers that are not square-free.

Note that  $\tilde{f}_1 * \dots * \tilde{f}_H(p^k) \leq (CH)^k$ . Thus, Shiu's bound (3.5) applied to the progression  $n \equiv A' \pmod{Wq}$  yields

$$\frac{Wq}{T} \sum_{\substack{n \leq T \\ n \equiv A' \pmod{Wq}}} \tilde{f}_1 * \dots * \tilde{f}_H(n) \ll \frac{W}{\phi(W)} \frac{1}{\log T} \exp \left( \sum_{w(N) < p < T} \frac{|f(p)|}{p} \right).$$

The latter expression equals  $\asymp S_{|f|}(N; Wq, A')$  by assumption, which completes the proof.  $\square$

4.  $W$ -TRICK

In the same way as the bounds mentioned in the introduction only apply to Fourier coefficients  $\frac{1}{x} \sum_{n \leq x} f(n) e(\alpha n)$  at an irrational phase  $\alpha$ , it is the case that an arbitrary multiplicative function  $f$  will most likely *correlate* with a given nilsequence, unless this sequence itself is sufficiently equidistributed. Thus, statements of the form

$$\frac{1}{N} \sum_{n \leq N} h(n) F(g(n)\Gamma) = o_{G/\Gamma}(1)$$

with  $h = f$  or  $h = f - S_f(N; 1, 1)$  cannot be expected to hold in general. On the other hand, it turns out to be sufficient to ensure that  $h$  is equidistributed in progressions to small moduli in order to resolve this problem. For arithmetic applications such as establishing a result of the form (1.8), this can be achieved with the help of the  $W$ -trick from [11]. The basic idea is to decompose  $f$  into a sum of functions that are equidistributed in progressions to small moduli. This is achieved by decomposing the range  $\{1, \dots, N\}$  into subprogressions modulo a product  $W(N)$  of small primes, which has the effect of fixing or eliminating the contribution from small primes on each of the subprogressions.

For multiplicative functions some minor modifications are necessary. Our aim is to decompose the interval  $\{1, \dots, N\}$  into subprogressions  $r \pmod{q}$  in such a way that

$$S_f(N; q, r) = (1 + o(1)) S_f(N; qq', r + qr')$$

for small  $q'$ . Thus,  $f$  should essentially have a constant average value when decomposing one of the given subprogressions into further subprogressions of small moduli  $q'$ . The example of the characteristic function of sums of two squares shows that we cannot in general choose  $q$  to be a product of small primes (consider the case where  $r \equiv 1 \pmod{2}$ ,  $q' = 2$  and  $r + qr' \equiv 3 \pmod{4}$ ), but rather need to allow  $q$  to be a product of small prime powers.

Recall that an integer is said to be  $k$ -smooth if it is free from prime factors exceeding  $k$ . Furthermore, recall from Section 1 that  $w(x) \leq \log \log x$  for all sufficiently large  $x$ , and that  $W(x) = \prod_{p \leq w(x)} p$ . Thus,  $\log W(x) = \sum_{p \leq w(x)} \log p \sim w(x)$  and hence

$$W(x) \leq (\log x)^{1+o(1)}.$$

Let  $q^* : \mathbb{N} \rightarrow \mathbb{N}$  be a function that satisfies  $q^*(x) \leq (\log x)^H$  for all sufficiently large integers  $x > 1$  and such that  $q^*(x)$  is a  $w(x)$ -smooth integer for every  $x \in \mathbb{N}$ . Let

$$\widetilde{W}(x) = q^*(x) W(x).$$

The factor  $q^*$  will allow us for instance to consider the representation function of sums of two squares or the representation function of a more general norm form with this  $W$ -trick.

We decompose the range  $[1, N]$  into subprogressions of the form

$$\left\{ 1 \leq m \leq N : m \equiv w_1 A \pmod{w_1 \widetilde{W}(N)} \right\},$$

where  $A \in (\mathbb{Z}/\widetilde{W}(N)\mathbb{Z})^*$  and where  $w_1 \geq 1$  is  $w(N)$ -smooth. Abbreviating  $\widetilde{W} = \widetilde{W}(N)$ , we have  $\gcd(w_1, \widetilde{W}n + A) = 1$  and hence  $f(w_1(\widetilde{W}n + A)) = f(w_1)f(\widetilde{W}n + A)$ . Thus, it

suffices to study the family of functions

$$\left\{ n \mapsto f(\widetilde{W}n + A) : \begin{array}{l} 0 < A < \widetilde{W}(N) \\ \gcd(A, \widetilde{W}) = 1 \end{array} \right\}.$$

Since large values of  $w_1$  form a sparse set, it is possible to bound their contribution to the sum  $\sum_{n \leq N} f(n)F(g(n)\Gamma)$  separately. In the remaining case,  $w_1$  is bounded above, which ensures that the range on which each function  $n \mapsto f(\widetilde{W}n + A)$  needs to be considered is always large. More precisely, combining the Cauchy–Schwarz inequality and a bound on the second moment of  $f$ , it is possible to discard the contribution from values of  $w_1$  such that  $v_p(w_1) > C_1 \log_p \log N$  for some prime  $p \leq w(N)$  and a sufficiently large constant  $C_1 > 1$ ; see [20, §3] for details. Thus for the purpose of arithmetic applications, it suffices to consider  $n \mapsto f(\widetilde{W}n + A)$  for  $n \in \{1, \dots, T\}$  with

$$T = \frac{N - Aw_1}{w_1 \widetilde{W}(N)} \gg N(\log N)^{-2H} \prod_{p < w(N)} (\log N)^{-C_1-1} \gg N(\log N)^{-2H-2C_1 \log \log N} \gg N^{1-o(1)}.$$

**Definition 4.1.** Let  $\mathcal{F}_H(E)$  denote the class of multiplicative functions from  $\mathcal{M}_H^*(E)$  with the additional property that for each  $f \in \mathcal{M}_H^*(E)$ , there is

- (1) a function  $\varphi : \mathbb{N} \rightarrow \mathbb{R}$  such that  $\varphi(x) = o(1)$  as  $x \rightarrow \infty$ , and
- (2) a function  $q^* : \mathbb{N} \rightarrow \mathbb{N}$  such that  $q^*(x)$  is  $w(x)$ -smooth and  $q^*(x) \leq (\log x)^H$  for all  $\varepsilon > 0$ ,

such that for all intervals  $I \subseteq \{1, \dots, x\}$  of length at least  $|I| > x(\log x)^{-E}$ , the estimate

$$\frac{q_0 \widetilde{W}(x)}{|I|} \sum_{\substack{m \in I \\ m \equiv A \pmod{q_0 \widetilde{W}(x)}}} f(m) = S_f(x; \widetilde{W}(x), A) + O\left(\varphi(x) S_{|f|}(x; \widetilde{W}(x), A)\right) \quad (4.1)$$

holds uniformly for all integers  $q_0$  such that  $0 < q_0 \leq (\log x)^E$  and for all representatives  $A \in \{1, \dots, q_0 \widetilde{W}(x)\}$  of invertible residues classes from  $(\mathbb{Z}/q_0 \widetilde{W}(x)\mathbb{Z})^*$ . Here  $\widetilde{W}(x) = q^*(x)W(x)$ , as before.

The estimate (4.1) will be referred to as ‘the major arc estimate’. We point out that a recent preprint of Irving [18] obtains strong results on the error term in (4.1) in the case where  $f$  is the divisor function. We will show in Section 5.2 that despite the restriction to invertible residues  $A \in (\mathbb{Z}/q_0 \widetilde{W}\mathbb{Z})^*$ , the estimate (4.1) implies that  $f(\widetilde{W}n + A) - S_f(x; \widetilde{W}, A)$  does not correlate with periodic sequences of period at most  $(\log x)^E$ . This information will be used in combination with a factorisation theorem to reduce the task of proving non-correlation for  $(f(\widetilde{W}n + A) - S_f(x; \widetilde{W}, A))$  with general nilsequences to the case where the nilsequence enjoys certain equidistribution properties and the Lipschitz function satisfies, in particular,  $\int_{G/\Gamma} F = 0$ .

## 5. THE NON-CORRELATION RESULT

This section contains a precise statement of the main result, which, informally speaking, shows the following: for every multiplicative function  $f \in \mathcal{F}_H(E)$ , for every invertible

residue  $A \in (\mathbb{Z}/\widetilde{W}(N)\mathbb{Z})^*$  and for parameters  $N$  and  $T$  such that  $N^{1-o(1)} \ll T \ll N$ , the sequence  $(f(\widetilde{W}n + A) - S_f(N; \widetilde{W}, A))_{n \leq T}$  is orthogonal to polynomial nilsequences provided  $E$  is sufficiently large with respect to  $H$ . In Section 5.2 we carry out a standard reduction of the main result to an equidistributed version, modelled on [14, §2].

**5.1. Statement of the main result.** Let  $G$  be a connected, simply connected,  $s$ -step nilpotent Lie group and let  $g : \mathbb{Z} \rightarrow G$  be a polynomial sequence on  $G$  as defined in [13, Definition 1.8]. Let  $\Gamma < G$  be a discrete cocompact subgroup. Then the compact quotient  $G/\Gamma$  is called a nilmanifold. Any Mal'cev basis  $\mathcal{X}$  (see [13, §2] for a definition) for  $G/\Gamma$  gives rise to a metric  $d_{\mathcal{X}}$  on  $G/\Gamma$  as described in [13, Definition 2.2]. This metric allows us to define Lipschitz functions on  $G/\Gamma$  as the set of functions  $F : G/\Gamma \rightarrow \mathbb{C}$  for which the Lipschitz norm (cf. [13, Definition 1.2])

$$\|F\|_{\text{Lip}} = \|F\|_{\infty} + \sup_{x, y \in G/\Gamma} \frac{|F(x) - F(y)|}{d_{\mathcal{X}}(x, y)}$$

is finite. If  $F$  is a 1-bounded Lipschitz function, then  $(F(g(n)\Gamma))_{n \in \mathbb{Z}}$  is called a (polynomial) nilsequence.

We are now ready to state the main result:

**Theorem 5.1.** *Let  $E, H$  be positive integers and suppose that  $f \in \mathcal{F}_H(E)$ . Let  $N$  and  $A$  be positive integer parameters with the property that  $0 < A < \widetilde{W}(N)$  and  $\gcd(W(N), A) = 1$ . Suppose further that  $T$  satisfies  $N^{1-o(1)} \ll T \ll N$  and that  $T, N > e^e$ . Let  $G/\Gamma$  be an  $s$ -step nilmanifold of positive dimension, let  $G_{\bullet}$  be a filtration of  $G$  and  $g \in \text{poly}(\mathbb{Z}, G_{\bullet})$  a polynomial sequence. Suppose that  $G/\Gamma$  has a  $Q_0$ -rational Mal'cev basis adapted to  $G_{\bullet}$  for some  $Q_0 \geq 2$  and let  $G/\Gamma$  be endowed with the metric defined by this basis. Let  $F : G/\Gamma \rightarrow \mathbb{C}$  be a 1-bounded Lipschitz function. Then, provided  $E \geq 1$  is sufficiently large with respect to  $d, m_G, H$  and  $1/\theta_f$ , we have*

$$\begin{aligned} \left| \frac{\widetilde{W}}{T} \sum_{n \leq T/\widetilde{W}} (f(\widetilde{W}n + A) - S_f(N; \widetilde{W}, A)) F(g(n)\Gamma) \right| &\ll (1 + \|F\|_{\text{Lip}}) \left\{ \right. \\ &Q_0^{O_{d, m_G}(1)} \frac{1}{(\log T)(\log \log T)^{1/(2s+4 \dim G)}} \exp \left( \sum_{w(N) < p \leq N} \frac{|f_1(p)| + \cdots + |f_t(p)|}{p} \right) \\ &+ \frac{(\log \log T)^{2H+1}}{\log T} \max_{1 \leq i \leq t} \exp \left( \sum_{w(N) < p < T} \frac{|f_1(p) + \cdots + f_t(p) - f_i(p)|}{p} \right) \\ &+ \mathbf{1}_{t>1} \frac{(\log \log T)^{2H+1}}{\log T} \max_{1 \leq i \leq t} \exp \left( \sum_{w(N) < p \leq N} \frac{|f_i(p)|}{p} \right) \\ &\left. + \frac{1}{\log w(N)} \max_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} S_{|f|}(N; \widetilde{W}, A') + \varphi(N) S_{|f|}(N; \widetilde{W}, A) \right\}, \end{aligned} \quad (5.1)$$



where  $\varphi$  is given by (4.1). The implied constants may depend on  $H$  as well as on the step and dimension of  $G$  and on the degree of the filtration  $G_\bullet$ .

**5.2. Reduction of Theorem 5.1 to the equidistributed case.** Proceeding similarly as in §2 of [14], we will reduce Theorem 5.1 to a special case that involves only equidistributed polynomial sequences. The tool that makes this reduction work is the following factorisation theorem [13, Thm 1.19] due to Green and Tao:

**Lemma 5.2** (Factorisation lemma, Green–Tao [13]). *Let  $m$  and  $d$  be positive integers, and let  $M_0, N, B \geq 1$  be real numbers. Let  $G/\Gamma$  be an  $m$ -dimensional nilmanifold together with a filtration  $G_\bullet$  of degree  $d$  and suppose that  $\mathcal{X}$  is an  $M_0$ -rational Mal'cev basis adapted to  $G_\bullet$ . Let  $g \in \text{poly}(\mathbb{Z}, G_\bullet)$  be a polynomial sequence. Then there is an integer  $M$  such that  $M_0 \ll M \ll M_0^{O_{B,m,d}(1)}$ , a rational subgroup  $G' \subseteq G$ , a Mal'cev basis  $\mathcal{X}'$  for  $G'/\Gamma'$  in which each element is an  $M$ -rational combination of the elements of  $\mathcal{X}$ , and a decomposition  $g = \varepsilon g' \gamma$  into polynomial sequences  $\varepsilon, g', \gamma \in \text{poly}(\mathbb{Z}, G_\bullet)$  with the following properties:*

- (1)  $\varepsilon : \mathbb{Z} \rightarrow G$  is  $(M, N)$ -smooth<sup>1</sup>;
- (2)  $g' : \mathbb{Z} \rightarrow G'$  takes values in  $G'$  and the finite sequence  $(g'(n)\Gamma')_{n \leq T}$  is totally  $M^{-B}$ -equidistributed in  $(G'\Gamma/\Gamma, d_{\mathcal{X}'})$ ;
- (3)  $\gamma : \mathbb{Z} \rightarrow G$  is  $M$ -rational<sup>2</sup> and the sequence  $(\gamma(n)\Gamma)_{n \in \mathbb{Z}}$  is periodic with period at most  $M$ .

The following proposition handles the special case of Theorem 5.1 where the nilsequence is equidistributed.

**Proposition 5.3** (Non-correlation, equidistributed case). *Let  $E, H \geq 1$  be integers and suppose that  $f \in \mathcal{M}_H^*(E)$ . Let  $N$  and  $T$  be integer parameters satisfying  $N^{1-o(1)} \ll T \ll N$ . Let  $m, d > 0$  and let  $\delta = \delta(N) \in (0, 1/2)$  depend on  $N$  in such a way that*

$$\log N \leq \delta(N)^{-1} \leq (\log N)^E.$$

*Let  $G/\Gamma$  be an  $s$ -step nilmanifold of dimension  $m_G$  together with a filtration  $G_\bullet$  of degree  $d$ , and suppose that  $\mathcal{X}$  is a  $\frac{1}{\delta(N)}$ -rational Mal'cev basis adapted to  $G_\bullet$ . This basis gives rise to the metric  $d_{\mathcal{X}}$ . Let  $q, r$  and  $A$  be integers such that  $0 \leq r < q \leq (\log N)^E$  and  $0 \leq A < \widetilde{W}$ , where  $\widetilde{W} = \widetilde{W}(N)$ , and suppose that  $\widetilde{W}r + A \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*$ .*

*Then there is  $E_0 \geq 1$ , depending on  $d, m_G, H$  and on  $\theta_f$  from Definition 1.2, such that the following holds for every  $E \geq 1$  that is sufficiently large with respect to  $d, m_G$  and  $H$ .*

*Let  $g \in \text{poly}(\mathbb{Z}, G_\bullet)$  be any polynomial sequence such that the finite sequence*

$$(g(n)\Gamma)_{n \leq T/(\widetilde{W}q)}$$

<sup>1</sup>The notion of smoothness was defined in [13, Def. 1.18]. A sequence  $(\varepsilon(n))_{n \in \mathbb{Z}}$  is said to be  $(M, N)$ -smooth if both  $d_{\mathcal{X}}(\varepsilon(n), \text{id}_G) \leq N$  and  $d_{\mathcal{X}}(\varepsilon(n), \varepsilon(n-1)) \leq M/N$  hold for all  $1 \leq n \leq N$ .

<sup>2</sup>A sequence  $\gamma : \mathbb{Z} \rightarrow G$  is said to be  $M$ -rational if for each  $n$  there is  $0 < r_n \leq M$  such that  $(\gamma(n))^{r_n} \in \Gamma$ ; see [13, Def. 1.17].

is totally  $\delta(N)^{E_0}$ -equidistributed. Let  $F : G/\Gamma \rightarrow \mathbb{C}$  be any 1-bounded Lipschitz function such that  $\int_{G/\Gamma} F = 0$ , and let  $I \subset \{1, \dots, T/(q\widetilde{W})\}$  be any discrete interval of length at least  $T/(q\widetilde{W}(\log N)^E)$ . Then:

$$\left| \frac{\widetilde{W}q}{T} \sum_{n \in I} f(\widetilde{W}(qn + r) + A) F(g(n)\Gamma) \right| \ll_{d, m_G, s, E, H, \theta_f} (1 + \|F\|_{\text{Lip}}) \mathcal{N}', \quad (5.2)$$

where

$$\begin{aligned} \mathcal{N}' = & \left( (\log \log T)^{-1/(2^{s+2} \dim G)} + \frac{Q^{10^s \dim G}}{(\log \log T)^{1/2}} \right) \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|f_1| * \dots * |f_t|}(N; \widetilde{W}q, A') \\ & + \mathbf{1}_{t>1} (\log \log T)^{2H} \max_{1 \leq i \leq t} \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|f_i|}(T; \widetilde{W}q, A') \\ & + (\log \log T)^2 \max_{1 \leq i \leq t} \frac{Wq}{\phi(Wq)} \frac{1}{\log T} \exp \left( \sum_{w(N) < p < T} \frac{|f_1(p) + \dots + f_t(p) - f_i(p)|}{p} \right). \end{aligned}$$

*Proof of Theorem 5.1 assuming Proposition 5.3.* We loosely follow the strategy of [14, §2]. In view of the first error term from Theorem 5.1, we may assume that  $Q_0 \leq \log(N)$ , as the theorem holds trivially otherwise. This implies that  $\mathcal{X}$  is a  $\log(N)$ -rational Mal'cev basis. Applying the factorisation lemma from above with  $T$  replaced by  $T/\widetilde{W}$ , with  $Q_0 = \log(N)$ , and with a parameter  $B$  that will be determined in course of the proof (as parameter  $E_0$  in an application of Proposition 5.3), we obtain a factorisation of  $g$  as  $\varepsilon g' \gamma$  with properties (1)–(3) from Lemma 5.2. In particular, there is  $Q$  such that  $\log N \leq Q \leq (\log N)^{O_{B, m_G, d}(1)}$  and such that  $g'$  takes values in a  $Q$ -rational subgroup  $G'$  of  $G$  and is  $Q^{-B}$ -equidistributed in  $G'\Gamma/\Gamma$ . Since  $\gamma$  is periodic with some period  $a \leq Q$ , the function  $n \mapsto \gamma(an + b)$  is constant for every  $b$ , that is,  $\gamma$  is constant on every progression

$$P_{a,b} := \{n \in [1, T/\widetilde{W}] : n \equiv b \pmod{a}\}$$

where  $0 \leq b < a$ . Let  $\gamma_b$  denote the value  $\gamma$  takes on  $P_{a,b}$  and note that

$$|P_{a,b}| \geq T/(2a\widetilde{W}) \geq T/(2Q\widetilde{W}).$$

Let  $g'_{a,b} : \mathbb{Z} \rightarrow G'$  be defined via

$$g'_{a,b}(n) = g'(an + b).$$

Since  $(g'(n)\Gamma)_{n \leq T/\widetilde{W}}$  is totally  $Q^{-B}$ -equidistributed in  $G'\Gamma/\Gamma$ , it is clear that every finite subsequence  $(g'_{a,b}(n)\Gamma)_{n \leq T/(Ca\widetilde{W})}$  is  $Q^{-B/2}$ -equidistributed if  $a, b$  and  $C$  are such that both  $0 \leq b < a \leq Q$  and  $C > 0$  and, furthermore,  $Q^{B/2} > Ca$  hold.

Let  $M \geq 1$  be an integer that will be chosen later depending on  $s$  and  $\dim G$ . By splitting each progression  $P_{a,b}$  into  $\ll Q(\log \log N)^{1/M}$  pieces  $P_{a,b}^{(j)}$  of diameter bounded by  $\ll T/(Q\widetilde{W}(\log \log N)^{1/M})$ , we may also arrange for  $\varepsilon$  to be almost constant. More precisely, the fact that  $\varepsilon$  is  $(Q, T/\widetilde{W})$ -smooth implies that

$$d_{\mathcal{X}}(\varepsilon(n), \varepsilon(n')) \leq |n - n'| Q \widetilde{W} T^{-1} \ll (\log \log N)^{-1/M}$$

for all  $n, n' \leq T/\widetilde{W}$  with  $|n - n'| \ll T/(Q\widetilde{W}(\log \log N)^{1/M})$ . By choosing  $B$  sufficiently large, we may ensure that  $Q^{B/2} \geq Q \log \log N$  and, hence, that the equidistribution properties of  $g'_{a,b}$  are preserved on the new bounded diameter pieces of  $P_{a,b}$ . Let  $\mathcal{P}$  denote the collection of all progressions  $P_{a,b}^{(j)}$  in our decomposition.

Since  $F$  is a Lipschitz function and since  $d_{\mathcal{X}}$  is right-invariant (cf. [13, Appendix A]), we deduce that

$$\begin{aligned} |F(\varepsilon(n)g'(n)\gamma(n)) - F(\varepsilon(n')g'(n)\gamma(n))| &\leq (1 + \|F\|)d(\varepsilon(n), \varepsilon(n')) \\ &\ll (1 + \|F\|)(\log \log N)^{-1/M} \end{aligned} \quad (5.3)$$

for all  $n, n' \in P_{a,b}$  with  $|n - n'| \ll T/(Q\widetilde{W}(\log \log N)^{1/M})$ . Thus, this bound holds in particular for any  $n, n' \in P_{a,b}^{(j)}$ . Let us fix one element  $n_{b,j}$  for each progression  $P_{a,b}^{(j)}$ . As we will show next, it will be sufficient to bound the correlation

$$\begin{aligned} &\left| \sum_{n \in P_{a,b}^{(j)}} \left( f(\widetilde{W}n + A) - S_f(N; \widetilde{W}, A) \right) F(\varepsilon(n_{b,j})g'(n)\gamma_b)\Gamma \right| \\ &= \left| \sum_{n: (an+b) \in P_{a,b}^{(j)}} \left( f(\widetilde{W}(an+b) + A) - S_f(N; \widetilde{W}, A) \right) F(\varepsilon(n_{b,j})g'_{a,b}(n)\gamma_b)\Gamma \right| \end{aligned} \quad (5.4)$$

for each bounded diameter piece  $P_{a,b}^{(j)}$ . Indeed, the estimate (5.3), applied with  $n' = n_{b,j}$  to each bounded diameter piece, shows that the error term that incurs from this reduction satisfies

$$\begin{aligned} &\sum_{P_{a,b}^{(j)} \in \mathcal{P}} \sum_{n \in P_{a,b}^{(j)}} |f(\widetilde{W}n + A)| \left| F(\varepsilon(n)g'(n)\gamma(n)) - F(\varepsilon(n_{b,j})g'(n)\gamma_b) \right| \\ &\ll \frac{(1 + \|F\|)S_{|f|}(N; \widetilde{W}, A)}{(\log \log N)^{1/M}}. \end{aligned} \quad (5.5)$$

Since  $\frac{\log \log N}{\log \log \log N} < w(N) \leq \log \log N$ , we have

$$\frac{1}{(\log \log N)^{1/M}} \ll_{s, \dim G} \frac{1}{\log w(N)},$$

which implies that the error term (5.5) is acceptable in view of (5.1).

We aim to estimate the correlation (5.4) with the help of Proposition 5.3. This task will be carried out in four steps, the first of which will be to bound the contribution from non-invertible residues  $\widetilde{W}b + A \pmod{\widetilde{W}a}$  to which Proposition 5.3 does not apply. The following two steps deal with checking the various assumptions of Proposition 5.3, while the fourth contains the actual application of the proposition.

*Step 1: Non-invertible residues.* We seek to bound the contribution to (5.1) of all progressions  $P_{a,b}^{(j)} \in \mathcal{P}$  such that  $\gcd(\widetilde{W}b + A, \widetilde{W}a) > 1$ . Let  $a' = \prod_{p > w(N)} p^{v_p(a)}$ . Since

$\gcd(A, \widetilde{W}) = 1$ , it suffices to check whether  $b$  satisfies  $\gcd(\widetilde{W}b + A, a') > 1$ . Thus, the contribution we seek to bound satisfies

$$\begin{aligned}
& \frac{\widetilde{W}}{T'} \sum_{d|a', d>1} \sum_{\substack{b<a: \\ \gcd(\widetilde{W}b+A, q')=d}} \sum_{\substack{n<T'/\widetilde{W} \\ n\equiv b \pmod{q'}}} |f(\widetilde{W}n + A)| \\
& \leq \sum_{d|a', d>1} \sum_{\substack{b<a: \\ \gcd(\widetilde{W}b+A, q')=d}} \frac{|f(d)|}{a} S_{|f|} \left( \frac{T}{d}; \frac{\widetilde{W}a}{d}, \frac{\widetilde{W}b+A}{d} \right) \\
& \leq \sum_{d|a', d>1} \frac{|f(d)|}{a} \phi\left(\frac{a}{d}\right) S_{|f|} \left( \frac{T}{d}; \frac{\widetilde{W}a}{d}, \frac{\widetilde{W}b+A}{d} \right) \\
& \leq \sum_{d|a', d>1} \frac{|f(d)|}{d} \max_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} S_{|f|}(N; \widetilde{W}, A')(1 + \varphi(N)),
\end{aligned}$$

where we made use of (4.1) again. Note that  $a' \leq (\log N)^E$  and that its total number of prime factors satisfies  $\Omega(a') \leq \frac{E \log \log N}{\log w(N)}$ . Thus:

$$\begin{aligned}
\sum_{d|a', d>1} \frac{|f(d)|}{d} & \leq \prod_{p|a'} \left( 1 + \frac{H}{p} + \frac{H^2}{p^2} + \dots \right) - 1 \\
& \leq \prod_{p|a'} \left( 1 + \frac{H}{p} \right) \left( 1 + \frac{H^2}{p^2(1 - \frac{H}{p})} \right) - 1 \\
& \leq \left( 1 + \frac{1}{w(N)} \right) \prod_{p|a'} \left( 1 + \frac{H}{p} \right) - 1 \\
& \leq \left( 1 + \frac{1}{w(N)} \right) \prod_{w(N) < p < w(N) + \Omega(a')} \left( 1 + \frac{1}{p} \right)^H - 1 \\
& \ll \left( 1 + \frac{1}{w(N)} \right) \left( \frac{\log(w(N) + \Omega(a'))}{\log w(N)} \right)^H - 1 \\
& \ll \left( 1 + \frac{1}{w(N)} \right) \frac{(\log(w(N) + \Omega(a')))^H - (\log w(N))^H}{(\log w(N))^H} + \frac{1}{w(N)}.
\end{aligned}$$

Using the elementary identity  $A^H - B^H = (A - B)(A^{H-1} + A^{H-2}B + \dots + B^{H-1})$ , the above is bounded by

$$\ll \log \left( \frac{w(N) + \Omega(a')}{w(n)} \right) \frac{(\log(w(N) + \Omega(a')))^{H-1}}{(\log w(N))^H} + \frac{1}{w(N)},$$

which in turn is bounded by

$$\ll \log \left( 1 + \frac{\Omega(a')}{w(n)} \right) (\log w(N))^{-1} + \frac{1}{w(N)} \ll \frac{1}{\log w(N)},$$

since  $\Omega(a') \leq \frac{E \log \log N}{\log w(N)} \leq w(N)$  by assumption on the function  $w$ . Thus, the total contribution of non-invertible residues  $\widetilde{W}b + A \pmod{\widetilde{W}a}$  is at most

$$O \left( \frac{1}{\log w(N)} \max_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} S_{|f|}(N; \widetilde{W}, A') \right),$$

which has been taken care of in (5.1). This leaves us to considering the case where the value of  $b$  does not impose an obstruction to applying Proposition 5.3.

*Step 2: Checking the initial conditions of Proposition 5.3.* The central assumption of Proposition 5.3 concerns the equidistribution of the nilsequence it is applied to. To verify this assumption in the case of (5.4), it is necessary to show that the conjugated polynomial sequence  $h : \mathbb{Z} \rightarrow G$  defined via  $h(n) = \gamma_b^{-1} g'_{a,b}(n) \gamma_b$  is in fact a polynomial sequence and that it inherits the equidistribution properties of  $g'_{a,b}(n)$ . Both these questions have been addressed in [14, §2] in a way that allows us to refer back to. Let  $H = \gamma_b^{-1} G' \gamma_b$  and define  $H_\bullet = \gamma_b^{-1}(G')_\bullet \gamma_b$ . Let  $\Lambda = \Gamma \cap H$  and define  $F_{b,j} : H/\Lambda \rightarrow \mathbb{R}$  via

$$F_{b,j}(x\Lambda) = F(\varepsilon(n_{b,j}) \gamma_b x \Gamma).$$

Then  $h \in \text{poly}(\mathbb{Z}, H_\bullet)$  and the correlation (5.4) that we seek to bound takes the form

$$\left| \sum_{n: (an+b) \in P_{a,b}^{(j)}} \left( f(\widetilde{W}(an+b) + A) - S_f(N; \widetilde{W}, A) \right) F_{b,j}(h(n)\Lambda) \right|. \quad (5.6)$$

The ‘Claim’ from the end of [14, §2] guarantees the existence of a Mal’cev basis  $\mathscr{Y}$  for  $H/\Lambda$  adapted to  $H_\bullet$  such that each basis element  $Y_i$  is a  $Q^{O(1)}$ -rational combination of basis elements  $X_i$ . Thus there is  $C' = O(1)$  such that  $\mathscr{Y}$  is  $Q^{C'}$ -rational. Furthermore, it implies that there is  $c' > 0$ , depending only on the dimension of  $G$  and the degree of  $G_\bullet$ , such that whenever  $B$  is sufficiently large the sequence

$$(h(n)\Lambda)_{n \leq T/(a\widetilde{W})} \quad (5.7)$$

is totally  $Q^{-c'B/2+O(1)}$ -equidistributed in  $H/\Lambda$ , equipped with the metric  $d_{\mathscr{Y}}$  induced by  $\mathscr{Y}$ . Taking  $B$  sufficiently large, we may assume that the sequence (5.7) is totally  $Q^{-c'B/4}$ -equidistributed. Finally, the ‘Claim’ also provides the bound  $\|F_{b,j}\| \leq Q^{C''} \|F\|$  for some  $C'' = O(1)$ . This shows that all conditions of Proposition 5.3 are satisfied except for  $\int_{H/\Lambda} F_{b,j} = 0$ .

*Step 3: The final condition.* The final condition that needs to be satisfied in order for Proposition 5.3 to apply is  $\int_{H/\Lambda} F_{b,j} = 0$ . This is where the major arc condition (4.1) comes into play, which notably requires that  $\gcd(\widetilde{W}b + A, \widetilde{W}a) = 1$ .

In fact, one can subtract off the mean value  $\int_{H/\Lambda} F_{b,j}$  of  $F_{b,j}$  at the expense of an additional error term, provided one can show that  $f(\widetilde{W}n + A) - S_f(N; \widetilde{W}, A)$  does not correlate with the characteristic function  $\mathbf{1}_{P_{a,b}^{(j)}}$  of the corresponding progression  $P_{a,b}^{(j)}$ . Note that the common difference of  $P_{a,b}^{(j)}$  satisfies  $a \leq Q \ll (\log N)^{O_{d,m_G,B}(1)}$ , which is bounded by  $(\log N)^E$ , provided  $E$  is sufficiently large in terms of  $d$ ,  $m_G$  and  $B$ . Similarly, the length of  $P_{a,b}^{(j)}$  satisfies

$$|P_{a,b}^{(j)}| \geq T/(2aQ\widetilde{W}(\log \log N)^{1/M}) \gg T/(a\widetilde{W}(\log N)^E),$$

provided  $E$  is sufficiently large in terms of  $d$ ,  $m_G$  and  $B$ . Condition (4.1) guarantees for every interval  $I \subset \{1, \dots, T/\widetilde{W}\}$  of length  $|I| \gg T/(\log T)^E$  the uniform estimate

$$\frac{q'\widetilde{W}}{|I|} \sum_{\substack{m \in I \\ m \equiv \widetilde{W}r' + A \pmod{q'\widetilde{W}(x)}}} f(m) = S_f(x; \widetilde{W}, A) + O\left(\varphi(x)S_{|f|}(x; \widetilde{W}, A)\right),$$

valid for all positive integers  $q' < (\log T)^E$  and for all  $r'$  such that  $0 \leq r' \leq q'$  and such that  $\gcd(\widetilde{W}r' + A, q') = 1$ . Thus, assuming  $E$  is sufficiently large to satisfy both conditions mentioned above, (4.1) applies. Let  $\mu_{b,j} := \int_{H/\Lambda} F_{b,j}$  and note that  $\mu_{b,j} \ll 1$ . Then the error term incurred by replacing for each  $P_{a,b}^{(j)}$  with  $\gcd(\widetilde{W}b + A, \widetilde{W}a) = 1$  the factor  $F_{b,j}(h(n)\Lambda)$  in (5.6) by  $(F_{b,j}(h(n)\Lambda) - \mu_{b,j})$  is bounded as follows:

$$\begin{aligned} & \left| \frac{\widetilde{W}}{T} \sum_{\substack{P_{a,b}^{(j)} \in \mathcal{P} \\ \gcd(\widetilde{W}b + A, a) = 1}} \mu_{b,j} \sum_{n \in P_{a,b}^{(j)}} \left( f(\widetilde{W}n + A) - S_f(N; \widetilde{W}, A) \right) \right| \\ & \ll \frac{\widetilde{W}}{T} \sum_{P_{a,b}^{(j)} \in \mathcal{P}} |P_{a,b}^{(j)}| \varphi(N) S_{|f|}(N; \widetilde{W}, A) \ll \varphi(N) S_{|f|}(N; \widetilde{W}, A). \end{aligned}$$

This error term has been taken care of in (5.1).

*Step 4: Application of Proposition 5.3.* In view of the work carried out in Steps 1–3, we may now assume that  $\gcd(\widetilde{W}b + A, \widetilde{W}a) = 1$  and that  $\int_{H/\Lambda} F_{b,j} = 0$  holds, and apply Proposition 5.3 with:

- $g = h$ ,  $q = a$ ,  $I = \{n : an + b \in P_{a,b}^{(j)}\}$ ,
- with a function  $\delta : \mathbb{N} \rightarrow \mathbb{R}$  such that  $\delta(N) = Q^{-C'} (= Q_0^{O_{d,m_G,B}(1)})$ , which ensures that  $\mathcal{Y}$  is  $\frac{1}{\delta(N)}$ -rational,
- with  $E$  sufficiently large to ensure that  $Q^{C'} < (\log N)^E$ , which in particular means that  $E$  depends on  $B$ , and
- with  $E_0 = c'B/(4C') = O_{d,m_G}(B)$ , where  $B$  is chosen sufficiently large for  $E_0$  to meet the requirements of Proposition 5.3.

Since there are  $\ll aQ(\log \log N)^{1/M}$  intervals  $P_{a,b}^{(j)}$  in the decomposition  $\mathcal{P}$ , this yields the bound

$$\begin{aligned} & \sum_{P_{a,b}^{(j)} \in \mathcal{P}} \left| \sum_{n: (an+b) \in P_{a,b}^{(j)}} \left( f(\widetilde{W}(an+b) + A) - S_f(N; \widetilde{W}, A) \right) F_{b,j}(h(n)\Lambda) \right| \\ & \ll aQ(\log \log N)^{1/M} (1 + Q^{O(1)} \|F\|) \frac{T}{\widetilde{W}a} \mathcal{N}'' \\ & \ll (1 + \|F\|) Q^{O(1)} (\log \log N)^{1/M} \frac{T}{\widetilde{W}} \mathcal{N}'', \end{aligned}$$

where

$$\begin{aligned} \mathcal{N}'' &= \left( (\log \log T)^{-1/(2^{s+2} \dim G)} + \frac{Q^{O_s, \dim G(1)}}{(\log \log T)^{1/2}} \right) \max_{A' \in (\mathbb{Z}/\widetilde{W}a\mathbb{Z})^*} S_{|f_1| * \dots * |f_t|}(N; \widetilde{W}a, A') \\ &+ \mathbf{1}_{t>1} (\log \log T)^{2H} \max_{1 \leq i \leq t} \max_{A' \in (\mathbb{Z}/\widetilde{W}a\mathbb{Z})^*} S_{|f_i|}(T; \widetilde{W}a, A') \\ &+ (\log \log T)^2 \max_{1 \leq i \leq t} \frac{\widetilde{W}a}{\phi(\widetilde{W}a)} \frac{1}{\log T} \exp \left( \sum_{w(N) < p < T} \frac{|f_1(p) + \dots + f_t(p) - f_i(p)|}{p} \right). \end{aligned}$$

To finish the proof of Theorem 5.1 it remains to remove the dependence on  $a$  in this bound. To this end, recall (cf. Step 1) that  $a' = \prod_{p > w(N)} p^{v_p(a)}$  satisfies  $\Omega(a') \leq E \log \log N$ , which implies

$$\frac{\widetilde{W}a}{\phi(\widetilde{W}a)} \ll \log(w(N) + E \log \log N) \ll_E \log \log \log N.$$

Invoking Shiu's lemma (Lemma 3.3) we furthermore obtain

$$S_{|f_1| * \dots * |f_t|}(N; \widetilde{W}a, A') \ll_E \frac{\log \log \log N}{\log N} \exp \left( \sum_{w(N) < p < T} \frac{|f_1(p)| + \dots + |f_t(p)|}{p} \right)$$

and

$$S_{|f_i|}(T; \widetilde{W}a, A') \ll_E \frac{\log \log \log N}{\log N} \exp \left( \sum_{w(N) < p < T} \frac{|f_i(p)|}{p} \right)$$

Inserting these bounds into the expression for  $\mathcal{N}''$  above and setting  $M = 2^{s+3} \dim G$ , concludes in view of (5.1) the deduction of Theorem 5.1.  $\square$

It remains to establish Proposition 5.3.

## 6. LINEAR SUBSEQUENCES OF EQUIDISTRIBUTED NILSEQUENCES

Our aim in this section is to study the equidistribution properties of families

$$\left\{ (g(Dn + D')\Gamma)_{n \leq T/D} : D \in [K, 2K) \right\}$$

of linear subsequences of an equidistributed sequence  $(g(n)\Gamma)_{n \leq T}$ , where  $D$  runs through dyadic intervals  $[K, 2K)$  for  $K \leq T^{1-1/t}$ .

We begin by recalling some essential definitions and notation. Let  $P : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  be a polynomial of degree at most  $d$  and let  $\alpha_0, \dots, \alpha_d \in \mathbb{R}/\mathbb{Z}$  be defined via

$$P(n) = \alpha_0 + \alpha_1 \binom{n}{1} + \dots + \alpha_d \binom{n}{d}.$$

Then the *smoothness norm* of  $g$  with respect to  $T$  is defined (c.f. Green–Tao [13, Def. 2.7]) as

$$\|P\|_{C^\infty[T]} = \sup_{1 \leq j \leq d} T^j \|\alpha_j\|_{\mathbb{R}/\mathbb{Z}}.$$

If  $\beta_0, \dots, \beta_d \in \mathbb{R}/\mathbb{Z}$  are defined via

$$P(n) = \beta_d n^d + \dots + \beta_1 n + \beta_0,$$

then (cf. [21, equation (14.3)]) the smoothness norm is bounded above by a similar expression in terms of the  $\beta_i$ , namely

$$\|P\|_{C^\infty[T]} \ll_d \sup_{1 \leq j \leq d} T^j \|j! \beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll_d \sup_{1 \leq j \leq d} T^j \|\beta_j\|_{\mathbb{R}/\mathbb{Z}}. \quad (6.1)$$

On the other hand, [13, Lemma 3.2] shows that there is a positive integer  $q \ll_d 1$  such that

$$\|q\beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll T^{-j} \|P\|_{C^\infty[T]}.$$

Apart from smoothness norms, we also require the notion of a horizontal character as defined in [13, Definition 1.5]. A continuous additive homomorphism  $\eta : G \rightarrow \mathbb{R}/\mathbb{Z}$  is called a *horizontal character* if it annihilates  $\Gamma$ . In order to formulate quantitative results, one defines a height function  $|\eta|$  for these characters. A definition of this height, called the *modulus* of  $\eta$ , may be found in [13, Definition 2.6]. All that we require to know about these heights is that there are at most  $M^{O(1)}$  horizontal characters  $\eta : G \rightarrow \mathbb{R}/\mathbb{Z}$  of modulus  $|\eta| \leq M$ .

The interest in smoothness norms and horizontal characters lies in Green and Tao’s ‘quantitative Leibman Theorem’:

**Proposition 6.1** (Green–Tao, Theorem 2.9 of [13]). *Let  $m_G$  and  $d$  be non-negative integers, let  $0 < \delta < 1/2$  and let  $N \geq 1$ . Suppose that  $G/\Gamma$  is an  $m_G$ -dimensional nilmanifold together with a filtration  $G_\bullet$  of degree  $d$  and that  $\mathcal{X}$  is a  $\frac{1}{\delta}$ -rational Mal’cev basis adapted to  $G_\bullet$ . Suppose that  $g \in \text{poly}(\mathbb{Z}, G_\bullet)$ . If  $(g(n)\Gamma)_{n \leq N}$  is not  $\delta$ -equidistributed, then there is a non-trivial horizontal character  $\eta$  with  $0 < |\eta| \ll \delta^{-O_{d,m_G}(1)}$  such that*

$$\|\eta \circ g\|_{C^\infty[N]} \ll \delta^{-O_{d,m_G}(1)}.$$

The following lemma shows that for polynomial sequences the notions of equidistribution and total equidistribution are equivalent with a polynomial dependence in the equidistribution parameter.



**Lemma 6.2.** *Let  $N$  and  $A$  be positive integers and let  $\delta : \mathbb{N} \rightarrow [0, 1]$  be a function that satisfies  $\delta(x)^{-t} \ll_t x$  for all  $t > 0$ . Suppose that  $G$  has a  $\frac{1}{\delta(N)}$ -rational Mal'cev basis adapted to the filtration  $G_\bullet$ . Suppose that  $g \in \text{poly}(\mathbb{Z}, G_\bullet)$  is a polynomial sequence such that  $(g(n)\Gamma)_{n \leq N}$  is  $\delta(N)^A$ -equidistributed. Then there is  $1 \leq B \ll_{d, m_G} 1$  such that  $(g(n)\Gamma)_{n \leq N}$  is totally  $\delta(N)^{A/B}$ -equidistributed, provided  $A/B > 1$  and provided  $N$  is sufficiently large.*

**Remark 6.3.** The Green–Tao factorisation theorem (cf. property (3) of Lemma 5.2) usually allows one to arrange for  $A > B$  to hold.

*Proof.* We allow all implied constants to depend on  $d$  and  $m_G$ . Let  $B \geq 1$  and suppose that  $(g(n)\Gamma)_{n \leq N}$  fails to be totally  $\delta(N)^{A/B}$ -equidistributed. Then there is a subprogression  $P = \{\ell n + b : 0 \leq n \leq m-1\}$  of  $\{1, \dots, N\}$  of length  $m > \delta(N)^{A/B} N$  such that the sequence  $(\tilde{g}(n))_{0 \leq n < m}$ , where  $\tilde{g}(n) = g(\ell n + b)$ , fails to be  $\delta(N)^{A/B}$ -equidistributed. Provided  $A > B$ , Proposition 6.1 implies that there is a non-trivial horizontal character  $\eta : G \rightarrow \mathbb{R}/\mathbb{Z}$  of modulus  $|\eta| < \delta(N)^{-O(A/B)}$  such that

$$\|\eta \circ \tilde{g}\|_{C^\infty[m]} \ll \delta(N)^{-O(A/B)}.$$

The lower bound on  $m$  implies that this is equivalent to the assertion

$$\|\eta \circ \tilde{g}\|_{C^\infty[N]} \ll \delta(N)^{-O(A/B)},$$

where we recall that the implied constant may depend on  $d$ .

Observing that  $\eta \circ g$  is a polynomial of degree at most  $d$ , let  $\eta \circ g(n) = \beta_d n^d + \dots + \beta_0$ . Then

$$\eta \circ \tilde{g}(n) = \sum_{i=0}^d n^i \sum_{j=i}^d \beta_j \binom{j}{i} \ell^i b^{j-i},$$

and, hence,

$$\sup_{1 \leq i \leq d} N^i \left\| \sum_{j=i}^d \beta_j \binom{j}{i} \ell^i b^{j-i} \right\| \ll \delta(N)^{-O(A/B)}.$$

This yields the bound

$$\left\| \sum_{j=i}^d \beta_j \binom{j}{i} \ell^i b^{j-i} \right\| \ll N^{-i} \delta(N)^{-O(A/B)} \quad (6.2)$$

for  $1 \leq i \leq d$ . Note that the lower bound on  $m$  implies that  $\ell < \delta(N)^{-A/B}$ . Using a downwards induction argument, we aim to show that

$$\|\ell^d \beta_j\| \ll N^{-j} \delta(N)^{-O(A/B)} \quad (6.3)$$

for all  $1 \leq j \leq d$ . For  $j = d$ , this is clear from the above. Suppose (6.3) holds for all  $j > i$ . For each  $i < j$  we then, in particular, have that

$$\left\| \ell^d \beta_j \binom{j}{i} b^{j-i} \right\| \ll_d \|\ell^d \beta_j\| b^{j-i} \ll_d N^{-j} \delta(N)^{-O(A/B)} b^{j-i} \ll_d N^{-i} \delta(N)^{-O(A/B)}.$$

Using the fact that  $\delta(N)^{-t} \ll_t N$  for all  $t > 0$ , we deduce that (6.3) holds for  $j = i$  from the above bounds and from (6.2). This shows that there is a non-trivial horizontal character, namely  $\ell^d \eta$ , of modulus at most  $\delta(N)^{-O(A/B)}$ , such that

$$\|\ell^d \eta \circ g\|_{C^\infty[N]} \ll \sup_{1 \leq i \leq d} N^i \|\ell^d \beta_i\|_{\mathbb{R}/\mathbb{Z}} \ll \delta(N)^{-O(A/B)},$$

where we made use of (6.1). Choosing  $B$  sufficiently large in terms of  $m$  and  $d$ , [21, Proposition 14.2(b)] implies that  $g$  is not  $\delta(N)^A$ -equidistributed, which is a contradiction.  $\square$

We are now ready to address the equidistribution properties of linear subsequences.

**Proposition 6.4.** *Let  $N$ ,  $T$  and  $W(N)$  be as before and let  $E_1 \geq 1$ . Let  $(A_D)_{D \in \mathbb{N}}$  be a sequence of integers such that  $|A_D| \leq D$  for every  $D \in \mathbb{N}$ . Further, let  $\delta : \mathbb{N} \rightarrow (0, 1)$  be a function that satisfies  $\delta(x)^{-t} \ll_t x$  for all  $t > 0$ . Suppose  $G/\Gamma$  has a  $\frac{1}{\delta(N)}$ -rational Mal'cev basis adapted to a filtration  $G_\bullet$  of degree  $d$ . Let  $g \in \text{poly}(G_\bullet, \mathbb{Z})$  be a polynomial sequence and suppose that the finite sequence  $(g(n)\Gamma)_{n \leq T}$  is totally  $\delta(T)^{E_1}$ -equidistributed in  $G/\Gamma$ . Then there is a constant  $c_1 \in (0, 1)$ , depending only on  $d$  and  $m_G$ , such that the following assertion holds for all integers*

$$K \in [(\log T)^{\log \log T}, T^{1-1/t}],$$

provided  $c_1 E_1 \geq 1$ .

Write  $g_D(n) = g(Dn + A_D)$  and let  $\mathcal{B}_K$  denote the set of integers  $D \in [K, 2K)$  for which

$$(g_D(n)\Gamma)_{n \leq T/D}$$

fails to be  $\delta(T)^{c_1 E_1}$ -equidistributed. Then

$$\#\mathcal{B}_K \ll K \delta(T)^{c_1 E_1}.$$

*Proof.* Let  $K \in [(\log T)^{\log \log T}, T^{1-1/t}]$  be a fixed integer and let  $c_1 > 0$  to be determined in the course of the proof. Suppose that  $E_1 > 1/c_1$ . Lemma 6.2 implies that for every  $D \in \mathcal{B}_K$ , the sequence  $(g_D(n)\Gamma)_{n \leq T/D}$  fails to be  $\delta(T)^{c_1 E_1 B}$ -equidistributed on  $G/\Gamma$  for some  $B > 0$  only depending on  $d$  and  $m_G$ . We continue to allow implied constants to depend on  $d$  and  $m_G$ . By Proposition 6.1, there is a non-trivial horizontal character  $\eta_D : G \rightarrow \mathbb{R}/\mathbb{Z}$  of magnitude  $|\eta_D| \ll \delta(T)^{-O(c_1 E_1)}$  such that

$$\|\eta_D \circ g_D\|_{C^\infty[T/D]} \ll \delta(T)^{-O(c_1 E_1)}. \quad (6.4)$$

For each non-trivial horizontal character  $\eta \rightarrow \mathbb{R}/\mathbb{Z}$  we define the set

$$\mathcal{D}_\eta = \{D \in \mathcal{B}_K : \eta_D = \eta\}.$$

Note that this set is empty unless  $|\eta| \ll \delta(T)^{-O(c_1 E_1)}$ . Suppose that

$$\#\mathcal{B}_K \geq K \delta(T)^{c_1 E_1}.$$

By the pigeon hole principle, there is some  $\eta$  of modulus  $|\eta| \ll \delta(T)^{-O(c_1 E_1)}$  such that

$$\#\mathcal{D}_\eta \geq K \delta(T)^{O(c_1 E_1)}.$$

Suppose

$$\eta \circ g(n) = \beta_d n^d + \dots \beta_1 n + \beta_0$$

and let

$$\eta \circ g_D(n) = \alpha_d^{(D)} n^d + \dots + \alpha_1^{(D)} n + \alpha_0^{(D)}$$

for any  $D \in \mathcal{B}_K$ . The quantities  $\alpha_j^{(D)}$  and  $\beta_j$  are linked through the relation

$$\alpha_j^{(D)} = D^j \sum_{i=j}^d \binom{j}{i} A_D^{i-j} \beta_i \quad (6.5)$$

for each  $1 \leq j \leq d$ . Thus, the bound (6.4) on the smoothness norm asserts that

$$\sup_{1 \leq j \leq d} \frac{T^j}{K^j} \|\alpha_j^{(D)}\| \ll \delta(T)^{-O(c_1 E_1)}. \quad (6.6)$$

With a downwards induction we deduce from (6.6) and (6.5) that

$$\sup_{1 \leq j \leq d} \frac{T^j}{K^j} \|D^j \beta_j\| \ll \delta(T)^{-O(c_1 E_1)}. \quad (6.7)$$

The bound (6.7) provides information on rational approximations of  $D^j \beta_j$  for many values of  $D$ . Our next aim is to use this information in order to deduce information on rational approximations of the  $\beta_j$  themselves. To achieve this, we employ the Waring trick that appeared in the *Type I* sums analysis in [14, §3], and begin by recalling the two lemmas that this trick rests upon. The first one is a recurrence result, [13, Lemma 3.2].

**Lemma 6.5** (Green–Tao [13]). *Let  $\alpha \in \mathbb{R}$ ,  $0 < \delta < 1/2$  and  $0 < \sigma < \delta/2$ , and let  $I \subseteq \mathbb{R}/\mathbb{Z}$  be an interval of length  $\sigma$  such that  $\alpha n \in I$  for at least  $\delta N$  values of  $n$ ,  $1 \leq n \leq N$ . Then there is some  $k \in \mathbb{Z}$  with  $0 < |k| \ll \delta^{-O(1)}$  such that  $\|k\alpha\| \ll \sigma \delta^{-O(1)}/N$ .*

The second, [14, Lemma 3.3], is a consequence of the asymptotic formula in Waring’s problem.

**Lemma 6.6** (Green–Tao [14]). *Let  $K \geq 1$  be an integer, and suppose that  $S \subseteq \{1, \dots, K\}$  is a set of size  $\alpha K$ . Suppose that  $t \geq 2^j + 1$ . Then  $\gg_{j,t} \alpha^{2t} K^j$  integers in the interval  $[1, tK^j]$  can be written in the form  $k_1^j + \dots + k_t^j$ ,  $k_1, \dots, k_t \in S$ .*

Returning to the proof of Proposition 6.4, let us consider the set

$$\overline{\mathcal{D}}_j = \left\{ m \leq s(2K)^j : \begin{array}{l} m = D_1^j + \dots + D_s^j \\ D_1, \dots, D_s \in \mathcal{D}_\eta \end{array} \right\}$$

for some  $s \geq 2^j + 1$ . Each element  $m$  of this set satisfies

$$\|\beta_j m\| \ll \delta(T)^{-O(c_1)} (K/T)^j, \quad 1 \leq j \leq d, \quad (6.8)$$

in view of (6.7). Thus, Lemma 6.6 implies that there are

$$\#\overline{\mathcal{D}}_j \gg \delta(T)^{O(c_1 E_1)} K^j$$

elements in this set. In view of the restrictions on  $K$  and the assumptions on the function  $\delta(x)$ , the conditions of Lemma 6.5 (on  $\sigma$  and  $\delta$ ) are satisfied provided  $T$  is sufficiently large. We conclude that there is an integer  $k_j$  such that

$$1 \leq k_j \ll \delta(T)^{-O(c_1 E_1)}$$

and such that

$$\|k_j \beta_j\| \ll \delta(T)^{-O(c_1 E_1)} T^{-j}.$$

Thus

$$\beta_j = \frac{a_j}{\kappa_j} + \tilde{\beta}_j, \tag{6.9}$$

where  $\kappa_j | k_j$ ,  $\gcd(a_j, \kappa_j) = 1$  and

$$0 \leq \tilde{\beta}_j \ll \delta(T)^{-O(c_1 E_1)} T^{-j}.$$

Hence,

$$\|\kappa_j \beta_j\| \ll \delta(T)^{-O(c_1 E_1)} T^{-j}. \tag{6.10}$$

Let  $\kappa = \text{lcm}(\kappa_1, \dots, \kappa_d)$  and set  $\tilde{\eta} = \kappa \eta$ . We proceed as in [14, §3]: The above implies that

$$\|\tilde{\eta} \circ g(n)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta(T)^{-O(c_1 E_1)} n/T,$$

which is small provided  $n$  is not too large. Indeed, if  $T' = \delta(T)^{c_1 E_1 C} T$  for some sufficiently large constant  $C \geq 1$ , only depending on  $d$  and  $m_G$ , and if  $n \in \{1, \dots, T'\}$ , then

$$\|\tilde{\eta} \circ g(n)\|_{\mathbb{R}/\mathbb{Z}} \leq 1/10.$$

Let  $\chi : \mathbb{R}/\mathbb{Z} \rightarrow [-1, 1]$  be a function of bounded Lipschitz norm that equals 1 on  $[-\frac{1}{10}, \frac{1}{10}]$  and satisfies  $\int_{\mathbb{R}/\mathbb{Z}} \chi(t) dt = 0$ . Then, by setting  $F := \chi \circ \tilde{\eta}$ , we obtain a Lipschitz function  $F : G/\Gamma \rightarrow [-1, 1]$  that satisfies  $\int_{G/\Gamma} F = 0$  and  $\|F\|_{\text{Lip}} \ll \delta(T)^{-O(c_1 E_1)}$ . Choosing, finally,  $c_1$  sufficiently small, only depending on  $d$  and  $m_G$ , we may ensure that

$$\|F\|_{\text{Lip}} < \delta(T)^{-E_1}$$

and, moreover, that

$$T' > \delta(T)^{E_1} T.$$

This choice of  $T'$ ,  $F$  and  $c_1$  implies that

$$\left| \frac{1}{T'} \sum_{1 \leq n \leq T'} F(g(n)\Gamma) \right| = 1 > \delta(T)^{E_1} \|F\|_{\text{Lip}},$$

which contradicts the fact that  $(g(n)\Gamma)_{n \leq T}$  is totally  $\delta(T)^{E_1}$ -equidistributed. This completes the proof of the proposition.  $\square$

## 7. EQUIDISTRIBUTION OF PRODUCT NILSEQUENCES

In this section we prove, building on material and techniques from [14, §3], a result on the equidistribution of products of nilsequences which will allow us to perform applications of the Cauchy–Schwarz inequality in Section 8. The specific form of the result is adjusted to the requirements of Section 8.

We begin by introducing the product sequences we shall be interested in. Suppose  $g \in \text{poly}(G_\bullet, \mathbb{Z})$  is a polynomial sequence. This is equivalent to the assertion that there exists an integer  $k$ , elements  $a_1, \dots, a_k$  of  $G$ , and integral polynomials  $P_1, \dots, P_k \in \mathbb{Z}[X]$  such that

$$g(n) = a_1^{P_1(n)} a_2^{P_2(n)} \dots a_k^{P_k(n)}.$$

Then, for any pair of integers  $(m, m')$ , the sequence  $n \mapsto (g(mn), g(m'n)^{-1})$  is a polynomial sequence on  $G \times G$  that may be represented by

$$(g(mn), g(m'n)^{-1}) = \left( \prod_{i=1}^k (a_i, 1)^{P_i(mn)} \right) \left( \prod_{i=1}^k (1, a_i)^{P_i(m'n)} \right)^{-1}.$$

The horizontal torus of  $G \times G$  arises as the direct product  $G/\Gamma[G, G] \times G/\Gamma[G, G]$  of horizontal tori for  $G$ . Let  $\pi : G \rightarrow G/\Gamma[G, G]$  be the natural projection map. Any horizontal character on  $G \times G$  restricts to a horizontal character on each of its factors. Thus, it takes the form  $\eta \oplus \eta'(g_1, g_2) := \eta(g_1) + \eta'(g_2)$  for horizontal characters  $\eta, \eta'$  of  $G$ . The following proposition will be applied in the proof of Proposition 5.3 to sequences  $g = g_D$  for unexceptional  $D$  in the sense of Proposition (6.4).

**Proposition 7.1.** *Let  $N$ ,  $T$  and  $\widetilde{W}(N)$  be as before and let  $E_2 \geq 1$ . Let  $(\tilde{D}_m)_{m \in \mathbb{N}}$  be a sequence of integers satisfying  $|\tilde{D}_m| < m$  for every  $m \in \mathbb{N}$ . Further, let  $\delta : \mathbb{N} \rightarrow (0, 1)$  be a function that satisfies  $\delta(x)^{-t} \ll_t x$  for all  $t > 0$ . Suppose  $G/\Gamma$  has a  $\frac{1}{\delta(T)}$ -rational Mal'cev basis adapted to a filtration  $G_\bullet$  of degree  $d$ . Let  $P \subset \{1, \dots, T\}$  be a discrete interval of length at least  $\delta(T)T$ . Suppose  $F : G/\Gamma \rightarrow \mathbb{C}$  is a 1-bounded function of bounded Lipschitz norm  $\|F\|_{\text{Lip}}$  and suppose that  $\int_{G/\Gamma} F = 0$ . Let  $g \in \text{poly}(G_\bullet, \mathbb{Z})$  and suppose that the finite sequence  $(g(n)\Gamma)_{n \leq T}$  is totally  $\delta(T)^{E_2}$ -equidistributed in  $G/\Gamma$ . Then there is a constant  $c_2 \in (0, 1)$ , only depending on  $d$  and  $m_G := \dim G$ , such that the following assertion holds for all integers*

$$K \in \left[ \exp((\log \log T)^2), \exp\left(\frac{\log T - (\log T)^{1/U}}{t}\right) \right],$$

where  $1 \leq U \ll 1$ , provided  $c_2 E_2 \geq 1$ .

Let  $\mathcal{E}_K$  denote the set of integer pairs  $(m, m') \in (K, 2K]^2$  such that

$$I_{m, m'} = \left\{ n \in \mathbb{N} : \begin{array}{l} nm + \tilde{D}_m \in P, \\ nm' + \tilde{D}_{m'} \in P \end{array} \right\} > \delta(N)^{c_2 E_2} T/K$$

and such that

$$\left| \sum_{\substack{n \leq T/\max(m, m') \\ n \in I_{m, m'}}} F(g_D(mn + \tilde{D}_m)\Gamma) \overline{F(g_D(m'n + \tilde{D}_{m'})\Gamma)} \right| > \frac{(1 + \|F\|_{\text{Lip}})\delta(T)^{c_2 E_2} T}{K}$$

holds. Then,

$$\#\mathcal{E}_K < K^2 \delta(T)^{O(c_2 E_2)},$$

uniformly for all  $K$  as above.

**Remark 7.2.** The above proposition essentially continues to hold when the variables  $(m, m')$  are restricted to pairs of primes. Due to a suitable choice of a cut-off parameter,  $X$ , that appears in Section 8.3, we will not need this variant of the proposition (cf. Section 8.6) and only provide a very brief account of it at the very end of this section.

*Proof.* To begin with, we endow  $G/\Gamma \times G/\Gamma$  with a metric by setting

$$d((x, y), (x', y')) = d_{G/\Gamma}(x, x') + d_{G/\Gamma}(y, y').$$

Let  $\tilde{F} : G/\Gamma \times G/\Gamma \rightarrow \mathbb{C}$  be defined via  $\tilde{F}(\gamma, \gamma') = F(\gamma) \overline{F(\gamma')}$ . This is a Lipschitz function. Indeed, the fact that  $F$  and  $\overline{F}$  are 1-bounded Lipschitz functions allows us to deduce that  $\|\tilde{F}\|_{\text{Lip}} \leq \|F\|_{\text{Lip}}$ . Let  $g_{m, m'} : \mathbb{N} \rightarrow G \times G$  be the polynomial sequence defined by

$$g_{m, m'}(n) = (g(nm + \tilde{D}_m), g(nm' + \tilde{D}_{m'})).$$

Furthermore, we write  $\Gamma' = \Gamma \times \Gamma$ . Then  $\tilde{F}$  satisfies

$$\int_{G/\Gamma \times G/\Gamma} \tilde{F}(\gamma, \gamma') d(\gamma, \gamma') = \int_{G/\Gamma} F(\gamma) \int_{G/\Gamma} \overline{F(\gamma')} d\gamma' d\gamma = 0.$$

Now, suppose that

$$K \in \left[ \exp((\log \log T)^2), \exp\left(\frac{\log T - (\log T)^{1/U}}{t}\right) \right]$$

and that

$$\mathcal{E}_K \geq K^2 \delta(T)^{c_2 E_2}.$$

For each pair  $(m, m') \in \mathcal{E}_K$ , we have

$$\begin{aligned} & \sum_{\substack{n \leq T/\max(m, m') \\ n \in I_{m, m'}}} F(g(mn + \tilde{D}_m)\Gamma) \overline{F(g(m'n + \tilde{D}_{m'})\Gamma)} \\ &= \sum_{\substack{n \leq T/\max(m, m') \\ n \in I_{m, m'}}} \tilde{F}(g_{m, m'}(n)\Gamma) > \frac{(1 + \|F\|_{\text{Lip}})\delta(T)^{c_2 E_2} T}{K}. \end{aligned} \tag{7.1}$$

Thus for every pair  $(m, m') \in \mathcal{E}_K$  the corresponding sequence

$$(\tilde{F}(g_{m, m'}(n)\Gamma))_{n \leq T/\max(m, m')}$$

fails to be totally  $\delta(T)^{c_2 E_2}$ -equidistributed. Lemma 6.2 implies that this finite sequence also fails to be  $\delta(T)^{c_2 E_2 B}$ -equidistributed for some  $B \geq 1$  that only depends on  $d$  and  $m_G$ . All implied constants in the sequel will be allowed to depend on  $d$  and  $m_G$ , without explicit mentioning. By [13, Theorem 2.9]<sup>3</sup>, there is for each pair  $(m, m') \in \mathcal{E}_K$  a non-trivial horizontal character

$$\tilde{\eta}_{m,m'} = \eta_{m,m'} \oplus \eta'_{m,m'} : G \times G \rightarrow \mathbb{R}/\mathbb{Z}$$

of magnitude  $\ll \delta(T)^{-O(c_2 E_2)}$  such that

$$\|\tilde{\eta}_{m,m'} \circ \tilde{g}_{m,m'}\|_{C^\infty[T/\max(m,m')]} \ll \delta(T)^{-O(c_2 E_2)}. \quad (7.2)$$

Given any non-trivial horizontal character  $\tilde{\eta} : G \times G \rightarrow \mathbb{R}/\mathbb{Z}$ , we define the set

$$\mathcal{M}_\eta = \{(m, m') \in \mathcal{E}_K \mid \tilde{\eta}_{m,m'} = \tilde{\eta}\}.$$

This set is empty unless  $|\tilde{\eta}| \ll \delta(T)^{-O(c_2 E_2)}$ . Pigeonholing over all non-trivial  $\tilde{\eta}$  of modulus bounded by  $\delta(T)^{-O(c_2 E_2)}$ , we find that there is some  $\tilde{\eta}$  amongst them for which

$$\#\mathcal{M}_{\tilde{\eta}} > K^2 \delta(T)^{O(c_2 E_2)}.$$

Let us fix such a character  $\tilde{\eta} = \eta \oplus \eta'$  and suppose without loss of generality that the component  $\eta$  is non-trivial. Suppose

$$\tilde{\eta} \circ (g(n), g(n')) = (\alpha_d n^d + \alpha'_d n'^d) + \cdots + (\alpha_1 n + \alpha'_1 n') + (\alpha_0 + \alpha'_0)$$

and define for  $(m, m') \in \mathcal{E}_K$  the coefficients  $\alpha_j(m, m')$ ,  $1 \leq j \leq d$ , via

$$\tilde{\eta} \circ g_{m,m'}(n) = \alpha_d(m, m') n^d + \cdots + \alpha_1(m, m') n + \alpha_0(m, m').$$

Then the bound (7.2) on the smoothness norm asserts that

$$\sup_{1 \leq j \leq d} \frac{T^j}{K^j} \|\alpha_j(m, m')\| \ll \delta(T)^{-O(c_2 E_2)}. \quad (7.3)$$

Observe that each  $\alpha_j(m, m')$ ,  $1 \leq j \leq d$ , satisfies

$$\alpha_j(m, m') = \sum_{i=j}^d \binom{i}{j} \left( \tilde{D}_m^{i-j} \alpha_i m^j + \tilde{D}_{m'}^{i-j} \alpha'_i m'^j \right). \quad (7.4)$$

We now aim to show with a downwards induction starting from  $j = d$  that

$$\alpha_j = \frac{a_j}{\kappa_j} + \tilde{\alpha}_j, \quad (7.5)$$

where  $1 \leq \kappa_j \ll \delta(T)^{-O(c_2 E_2)}$ ,  $\gcd(a_j, \kappa_j) = 1$ , and

$$\tilde{\alpha}_j \ll \delta(T)^{-O(c_2 E_2)} T^{-j}. \quad (7.6)$$

Suppose  $j_0 \leq d$  and that the above holds for all  $j > j_0$ . Set  $k_{j_0} = \text{lcm}(\kappa_{j_0+1}, \dots, \kappa_d)$  if  $j_0 < d$ , and  $k_{j_0} = 1$  when  $j_0 = d$ . Note that  $k_{j_0} \ll \delta(T)^{-O(c_2 E_2)}$ .

Pigeonholing, we find that there is  $\tilde{m}'$  such that  $m' = \tilde{m}'$  for  $\gg K \delta(T)^{O(c_2 E_2)}$  pairs  $(m, m') \in \mathcal{M}_{\tilde{\eta}}$ . Amongst these there are furthermore  $\gg K \delta(T)^{O(c_2 E_2)}$  values of  $m$  that

<sup>3</sup>The  $Q$ -rational Mal'cev basis for  $G/\Gamma$  induces one for  $G/\Gamma \times G/\Gamma$ . Thus [13, Theorem 2.9] is applicable.

belong to the same fixed residue class modulo  $k_{j_0}$ . Denote this set of integers  $m$  by  $\mathcal{M}'$ . Suppose  $m = k_{j_0}m_1 + m_0 \in \mathcal{M}'$ . Letting  $\{x\}$  denote the fractional part of  $x \in \mathbb{R}$ , we then have

$$\{\tilde{D}_m^{i-j_0}\alpha_i m^{j_0}\} = \left\{ \tilde{D}_m^{i-j_0}\tilde{\alpha}_i m^{j_0} + \frac{a_i m_0^{j_0}}{\kappa_i} \right\}, \quad (i \geq j_0),$$

where, in view of (7.6),

$$\tilde{D}_m^{i-j_0}\tilde{\alpha}_i m^{j_0} \ll \delta(T)^{-O(c_2 E_2)} K^i T^{-i}.$$

Since  $m_0$  is fixed, it thus follows from (7.3), (7.4), (7.5) and the above bound that as  $m$  varies over  $\mathcal{M}'$ , the value of

$$\|\alpha_{j_0} m^{j_0}\|$$

lies in a fixed interval of length  $\ll \delta(T)^{-O(c_2 E_2)} K^{j_0} T^{-j_0}$ .

We aim to make use of this information in combination with the Waring trick from [14, §3] that was already employed in Section 6. For this purpose, we consider the set of integers

$$\mathcal{M}^* = \left\{ m \leq s(2K)^{j_0} : \begin{array}{l} m = m_1^{j_0} + \cdots + m_s^{j_0} \\ m_1, \dots, m_s \in \mathcal{M}' \end{array} \right\}$$

with  $s \geq 2^{j_0} + 1$ . For each element  $m \in \mathcal{M}^*$  of this set,  $\|\alpha_{j_0} m\|$  lies in an interval of length  $\ll_s \delta(T)^{-O(c_2 E_2)} K^{j_0} T^{-j_0}$ . Furthermore, Lemma 6.6 implies that  $\#\mathcal{M}^* \gg \delta(T)^{O(c_2 E_2)} K^{j_0}$ . The restrictions on the size of  $K$  and the assumptions on the function  $\delta$  imply that the conditions of Lemma 6.5 are satisfied once  $T$  is sufficiently large. Thus, assuming  $T$  is sufficiently large, there is an integer  $1 \leq \kappa'_{j_0} \ll \delta(T)^{-O(c_2 E_2)}$  such that

$$\|\kappa'_{j_0} \alpha_{j_0}\| \ll \delta(T)^{-O(c_2 E_2)} T^{-j_0},$$

i.e.

$$\alpha_{j_0} = \frac{a_{j_0}}{\kappa_{j_0}} + \tilde{\alpha}_{j_0},$$

where  $\kappa_{j_0} | \kappa'_{j_0}$ ,  $\gcd(a_{j_0}, \kappa_{j_0}) = 1$  and  $\tilde{\alpha}_{j_0} \ll \delta(T)^{-O(c_2 E_2)} T^{-j_0}$ , as claimed.

In particular, we have

$$\|\kappa_j \alpha_j\| \ll \delta(T)^{-O(c_2 E_2)} T^{-j} \tag{7.7}$$

for  $1 \leq j \leq d$ . Proceeding as in [14, §3], let  $\kappa = \text{lcm}(\kappa_1, \dots, \kappa_d)$  and set  $\tilde{\eta} = \kappa \eta$ . Then (7.7) implies that

$$\|\tilde{\eta} \circ g\|_{C^\infty[T]} = \sup_{1 \leq j \leq d} T^j \|\kappa \alpha_j\| \ll \delta(T)^{-O(c_2 E_2)},$$

which in turn shows that

$$\|\tilde{\eta} \circ g(n)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta(T)^{-O(c_2 E_2)} n/T$$

for every  $n \in \{1, \dots, T\}$ . Note that the latter bound can be controlled by restricting  $n$  to a smaller range. For this, set  $T' = \delta(T)^{c_2 E_2 C} T$  for some constant  $C \geq 1$  depending only on  $d$  and  $m_G$ , chosen sufficiently large to guarantee that

$$\|\tilde{\eta} \circ g(n)\|_{\mathbb{R}/\mathbb{Z}} \leq 1/10,$$



whenever  $n \in \{1, \dots, T'\}$ . Let  $\chi : \mathbb{R}/\mathbb{Z} \rightarrow [-1, 1]$  be a function of bounded Lipschitz norm that equals 1 on  $[-\frac{1}{10}, \frac{1}{10}]$  and satisfies  $\int_{\mathbb{R}/\mathbb{Z}} \chi(t) dt = 0$ . Then, by setting  $F := \chi \circ \tilde{\eta}$ , we obtain a function  $F : G/\Gamma \rightarrow [-1, 1]$  such that  $\int_{G/\Gamma} F = 0$  and  $\|F\|_{\text{Lip}} \ll \delta(T)^{-O(c_2 E_2)}$ . Choosing  $c_2$  sufficiently small, we may ensure that

$$\|F\|_{\text{Lip}} < \delta(T)^{-E_2}$$

and, moreover, that

$$T' > \delta(T)^{E_2} T.$$

The quantities  $T'$ ,  $F$  and  $c_2$  are chosen in such a way that

$$\left| \frac{1}{T'} \sum_{1 \leq n \leq T'} F(g(n)\Gamma) \right| = 1 > \delta(T)^{E_2} \|F\|_{\text{Lip}},$$

This contradicts the fact that  $(g(n)\Gamma)_{n \leq T}$  is totally  $\delta(T)^{E_2}$ -equidistributed and completes the proof of the proposition.  $\square$

**7.1. Restriction of Proposition 7.1 to pairs of primes.** We end this section by making the contents of Remark 7.2 more precise. The variables  $(m, m')$  in Proposition 7.1 can without much additional effort be restricted to range over pairs of primes. It is clear that in the above proof all applications of the pigeonhole principle that involve the parameters  $m$  and  $m'$  have to be restricted to the set of primes. The only true difference lies in the application of Waring's result: Lemma 6.6 needs to be replaced by the following one.

**Lemma 7.3.** *Let  $K \geq 1$  be an integer and let  $S \subset \{1, \dots, K\} \cap \mathcal{P}$  be a subset of the primes less than  $K$ . Suppose  $\#S = \alpha \frac{K}{\log K}$ . Let  $s \geq 2^k + 1$ . Let  $X \subset \{1, \dots, sK^k\}$  denote the set of integers that are representable as  $p_1^k + \dots + p_s^k$  with  $p_1, \dots, p_s \in S$ . Then*

$$|X| \gg_{k,s} \alpha^{2s} K^k,$$

as  $K \rightarrow \infty$ .

*Proof.* Let  $I_s(N)$  denote the number of solutions to the equation

$$p_1^k + \dots + p_s^k = N$$

in positive prime numbers  $p_1, \dots, p_s$ . Hua's asymptotic formula [17, Theorem 11] for the Waring–Goldbach problem implies that

$$I_s(N) \ll_{k,s} \frac{N^{s/k-1}}{(\log N)^s}.$$

Thus, for  $1 \leq n \leq sK^k$ , we have

$$I_s(n) \ll_{k,s} \frac{K^{s-k}}{(\log K)^s}.$$

Hence,

$$\alpha^{2s} \frac{K^{2s}}{(\log K)^{2s}} = \left( \sum_{n=1}^{sK^k} I_s(n) \right)^2 \leq |X| \sum_n I_s^2(n) \ll_{k,s} |X| K^k \frac{K^{2s-2k}}{(\log K)^{2s}} \ll_{k,s} |X| \frac{K^{2s-k}}{(\log K)^{2s}}.$$

Rearranging completes the proof of the lemma.  $\square$

## 8. PROOF OF PROPOSITION 5.3

In this section we prove Proposition 5.3 by invoking the possibly trivial Dirichlet decomposition from Definition 1.2. Let  $f \in \mathcal{F}_H(E)$  and let  $f = f_1 * \dots * f_t$  be a Dirichlet decomposition with the properties of Definition 1.2 and 1.6 for some  $t \leq H$ . We are given integers  $q$  and  $r$  such that  $0 \leq r < q \leq (\log N)^E$  and such that  $\widetilde{W}r + A \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*$ . Recall that  $g \in \text{poly}(\mathbb{Z}, G_\bullet)$  is a polynomial sequence with the property that  $(g(n)\Gamma)_{n \leq T/(\widetilde{W}q)}$  is totally  $\delta(N)^{E_0}$ -equidistributed in  $G/\Gamma$ . Let  $I \subset \{1, \dots, T/\widetilde{W}q\}$  be a discrete interval of length at least  $T/(\widetilde{W}q(\log N)^E)$ . Our aim is to bound above the expression

$$\frac{\widetilde{W}q}{T} \sum_{n \in I} f(\widetilde{W}(qn + r) + A) F(g(n)\Gamma).$$

**8.1. Reduction by hyperbola method.** Taking into account the identity  $f = f_1 * \dots * f_t$ , the correlation from Proposition 5.3 may be written as

$$\begin{aligned} & \frac{\widetilde{W}q}{T} \sum_{n \leq T/\widetilde{W}} \mathbf{1}_I(n) f(\widetilde{W}(qn + r) + A) F(g(n)\Gamma) \\ &= \frac{\widetilde{W}q}{N} \sum_{\substack{d_1 \dots d_t \leq T \\ d_1 \dots d_t \equiv \widetilde{W}r + A \\ (\text{mod } \widetilde{W}q)}} f_1(d_1) f_2(d_2) \dots f_t(d_t) F\left(g\left(\frac{d_1 \dots d_t - A - r\widetilde{W}}{\widetilde{W}q}\right) \Gamma\right) \mathbf{1}_P(d_1 \dots d_t), \end{aligned} \quad (8.1)$$

where  $P$  is the finite progression defined via  $P = \widetilde{W}(qI + r) + A$ . Our first step is to split this summation via inclusion-exclusion into a finite sum of weighted correlations of individual factors  $f_i$  with a nilsequence. To describe these weighted correlations, let  $i \in \{1, \dots, t\}$ . For every  $j \neq i$ , let  $d_j$  be a fixed positive integer and write  $D_i := \prod_{j \neq i} d_j$ . Let  $a_i \in [0, T/D_i]$  be an integer. Weighted correlations involving  $f_i$  will then take the form:

$$\begin{aligned} & \frac{\widetilde{W}}{T} \left( \prod_{j \neq i} f_j(d_j) \right) \sum_{\substack{a_i < d_i \leq T/D_i \\ d_i D_i \equiv A + r\widetilde{W} \pmod{\widetilde{W}q}}} \mathbf{1}_P(d_i D_i) f_i(d_i) F\left(g\left(\frac{d_i D_i - A - r\widetilde{W}}{\widetilde{W}q}\right) \Gamma\right) \\ &= \frac{\widetilde{W}}{T} \left( \prod_{j \neq i} f_j(d_j) \right) \sum_{\substack{a_i - D'_i < n \leq \frac{T - D'_i}{D_i \widetilde{W}q}}} f_i(\widetilde{W}qn + D'_i) F\left(g(D_i n + D''_i) \Gamma\right) \mathbf{1}_I(D_i n + D''_i), \end{aligned} \quad (8.2)$$

for suitable integers  $D'_i, D''_i$ , determined by the values of  $D_i \pmod{\widetilde{W}q}$  and  $A + r\widetilde{W}$ . In order to bound correlations of the form (8.2), we need to ensure that  $d_i$  runs over a sufficiently long range, which will be achieved by arranging for  $D_i \leq T^{1-1/t}$  to hold.

Let  $\tau = T^{1-1/t}$  and note that  $D_i = D_j \frac{d_j}{d_i}$ . Hence,

$$D_i > \tau \iff d_j > \frac{\tau d_i}{D_j}.$$

With the help of this equivalence, the function  $\mathbf{1} : \mathbb{Z}^t \rightarrow 1$  can be decomposed as follows. Suppose  $d_1 \dots d_t \leq T$ . Then

$$\begin{aligned} \mathbf{1}(d_1, \dots, d_t) &= \mathbf{1}_{D_1 \leq \tau} + \mathbf{1}_{D_1 > \tau} \left( \mathbf{1}_{D_2 \leq \tau} + \mathbf{1}_{D_2 > \tau} \left( \mathbf{1}_{D_3 \leq \tau} + \dots \left( \mathbf{1}_{D_t \leq \tau} + \mathbf{1}_{D_t > \tau} \right) \dots \right) \right) \\ &= \mathbf{1}_{D_1 \leq \tau} + \mathbf{1}_{D_1 > \tau} \left( \mathbf{1}_{D_2 \leq \tau} \mathbf{1}_{d_2 > \frac{\tau d_1}{D_2}} + \mathbf{1}_{D_2 > \tau} \left( \mathbf{1}_{D_3 \leq \tau} \mathbf{1}_{d_3 > \frac{\tau \max(d_1, d_2)}{D_3}} + \dots \right. \right. \\ &\quad \left. \left. \dots + \mathbf{1}_{D_{t-1} > \tau} \left( \mathbf{1}_{D_t \leq \tau} + \mathbf{1}_{D_t > \tau} \right) \dots \right) \right) \\ &= \mathbf{1}_{D_1 \leq \tau} + \mathbf{1}_{D_2 \leq \tau} \mathbf{1}_{d_2 > \frac{\tau d_1}{D_2}} + \mathbf{1}_{D_3 \leq \tau} \mathbf{1}_{d_3 > \frac{\tau \max(d_1, d_2)}{D_3}} + \dots + \mathbf{1}_{D_t \leq \tau} \mathbf{1}_{d_t > \frac{\tau \max(d_1, \dots, d_{t-1})}{D_t}}. \end{aligned}$$

Thus,

$$\sum_{d_1 \dots d_t < T} = \sum_{i=1}^t \sum_{D \leq T^{1-1/t}} \sum_{\substack{d_1, \dots, \widehat{d_i}, \dots, d_t \\ D_i = D}} \sum_{\substack{d_i \leq T/D_i \\ d_i > \tau \max(d_1, \dots, d_{i-1})/D_i}}.$$

This shows that the original summation (8.1) may be decomposed as a sum of summations of the shape (8.2) while only increasing the total number of terms by a factor of order  $O(t)$ . Expressing, if necessary, the summation range

$$(\tau \max(d_1, \dots, d_{i-1})/D_i, T/D_i)$$

of  $d_i$  as the difference of two intervals starting from 1, we can ensure that  $d_i$  runs over an interval of length  $\gg T/D_i \gg T^{1/t}$ . The correlation now decomposes as:

$$\begin{aligned} &\sum_{\substack{d_1 \dots d_t \leq T \\ d_1 \dots d_t \equiv \widetilde{W}r + A \\ (\text{mod } \widetilde{W}q)}} f_1(d_1) f_2(d_2) \dots f_t(d_t) F \left( g \left( \frac{d_1 \dots d_t - A - r\widetilde{W}}{\widetilde{W}q} \right) \Gamma \right) \mathbf{1}_P(d_1 \dots d_t) \\ &\leq \sum_{i=1}^t \sum_{k=1}^{\frac{1-1/t}{\log 2} \log T} \sum_{D \sim 2^k} \sum_{\substack{d_1, \dots, \widehat{d_i}, \dots, d_t \\ D_i = D}} \left( \prod_{j \neq i} |f_j(d_j)| \right) \\ &\quad \left| \sum_{\substack{n \leq T/D \\ n > \frac{\tau}{D} \max(d_1, \dots, d_{i-1}) \\ Dn + D'' \in I}} f_i(\widetilde{W}qn + D') F(g(Dn + D'')) \right|. \end{aligned} \tag{8.3}$$

Our next aim is to analyse the innermost sum of (8.3) as  $D \sim 2^k$  varies. Setting  $E_1 = E_0$ , we deduce from Proposition 6.4 that whenever  $2^k \in [\exp((\log \log T)^2), (\log T)^{1-1/t}]$  then

there is a set  $\mathcal{B}_{2^k}$  of cardinality at most  $O(\delta(N)^{c_1 E_0} 2^k)$  such that for each  $D \sim 2^k$  with  $D \notin \mathcal{B}_{2^k}$  the sequence

$$(g_D(n)\Gamma)_{n \leq T/\widetilde{W}_q}, \quad g_D(n) := g(Dn + D''),$$

is  $\delta(N)^{c_1 E_0}$ -equidistributed. Before turning to the case of  $D \notin \mathcal{B}_{2^k}$ , we bound the total contribution from exceptional  $D$ , that is, from  $D \in \mathcal{B}_{2^k}$  and from  $D \leq \exp((\log \log T)^2)$ .

**8.2. Contribution from exceptional  $D$ .** Let  $\mathcal{B}_{2^k}$  denote the exceptional set from the previous section.

**Lemma 8.1.** *Whenever  $E_0$  is sufficiently large in terms of  $d$ ,  $m_G$  and  $H$ , the following two estimates hold:*

$$\begin{aligned} & \sum_{\frac{(\log \log T)^2}{\log 2} < k \leq \frac{(1-1/t) \log T}{\log 2}} \sum_{D \in \mathcal{B}_{2^k}} \sum_{\substack{d_1 \dots d_t \leq T \\ d_1 \dots d_t \equiv \widetilde{W}r + A \pmod{\widetilde{W}q} \\ D_i = D}} |f_1(d_1)f_2(d_2) \dots f_t(d_t)| \mathbf{1}_P(d_1 \dots d_t) \\ & \ll_t \frac{T}{\widetilde{W}_q} \frac{1}{(\log T)^2}, \end{aligned}$$

and

$$\begin{aligned} & \sum_{\substack{D \leq \exp((\log \log T)^2) \\ \gcd(D, \widetilde{W}) = 1}} \sum_{\substack{d_1 \dots d_t \leq T \\ d_1 \dots d_t \equiv \widetilde{W}r + A \pmod{\widetilde{W}q} \\ D_i = D}} |f_1(d_1)f_2(d_2) \dots f_t(d_t)| \mathbf{1}_P(d_1 \dots d_t) \\ & \ll (\log \log T)^{2H} \frac{T}{\widetilde{W}_q} \max_{A' \in (\mathbb{Z}/\widetilde{W}_q\mathbb{Z})^*} S_{|f_i|}(T; \widetilde{W}_q, A'). \end{aligned}$$

**Remark.** Note that the above bounds only contribute to the bound in Proposition 5.3 in the case where the Dirichlet convolution is non-trivial, that is, when  $t > 1$ . In view of (1.6), the contribution from the first part is dominated by the contribution from the second part and may therefore be ignored. Note that the second part is accounted for in the bound from Proposition 5.3.

*Proof.* Set

$$\overline{f}_i(n) := |f_1 * \dots \widehat{f}_i \dots * f_t(n)|$$

and note that  $\overline{f}_i(n) \leq (CH)^{\Omega(n)}$  for some constant  $C$ . This implies a second moment bound of the form

$$\sum_{n \leq x} \overline{f}_i(n)^2 \leq x(\log x)^{O(H)}.$$

Similarly, we have

$$\sum_{n \leq x} |f_i(n)| \ll x(\log x)^{O(H)}.$$

Since Proposition 6.4 provides the bound  $\#\mathcal{B}_{2^k}^* \ll \delta(N)^{c_1 E_0} 2^k$ , a trivial application of the Cauchy–Schwarz inequality yields

$$\begin{aligned}
& \sum_{D \in \mathcal{B}_{2^k}^*} |f_1 * \dots \widehat{f_i} \dots * f_t(D)| \sum_{\substack{n \leq T/D \\ nD \equiv A + \widetilde{W}r \pmod{\widetilde{W}q} \\ nD \in P}} |f_i(n)| \\
&= \sum_{\substack{n \leq T/2^k \\ \gcd(n, W)=1}} |f_i(n)| \sum_{\substack{D \in \mathcal{B}_{2^k}^* \\ nD \equiv A + \widetilde{W}r \pmod{\widetilde{W}q} \\ nD \in P}} \overline{f_i}(D) \\
&\leq \sum_{\substack{n \leq T/2^k \\ \gcd(n, W)=1}} |f_i(n)| 2^k \delta(N)^{c_1 E_0} k^{O(H)} \\
&\leq T(\log T)^{O(H)} \delta(N)^{c_1 E_0} k^{O(H)}.
\end{aligned}$$

Recall that  $c_1$  only depends on  $d$  and  $m_G$ , and that by assumption of Proposition 5.3 we have  $\delta(N) \ll (\log T)^{-1}$ . Thus, the first part of the lemma follows by taking into account that the sum in  $k$  has length at most  $\log T$  and  $k^{O(H)} < (\log T)^{O(H)}$  and by choosing  $E_0$  sufficiently large in terms of  $d$ ,  $m_G$  and  $H$ .

Concerning the small values of  $D$ , we have

$$\begin{aligned}
& \sum_{\substack{D \leq \exp((\log \log T)^2) \\ \gcd(D, W)=1}} \overline{f_i}(D) \sum_{\substack{n \leq T/D \\ nD \equiv A + \widetilde{W}r \pmod{\widetilde{W}q} \\ nD \in P}} |f_i(n)| \\
&\ll \sum_{\substack{D \leq \exp((\log \log T)^2) \\ \gcd(D, W)=1}} \overline{f_i}(D) \frac{T}{D\widetilde{W}q} S_{|f_i|}(T/D; \widetilde{W}q, \overline{D}(A + \widetilde{W}r)),
\end{aligned}$$

where  $\overline{D}D \equiv 1 \pmod{\widetilde{W}q}$ . By property (3) of Definition 1.2, this in turn is bounded by

$$\ll \frac{T}{\widetilde{W}q} \sum_{\substack{D \leq \exp((\log \log T)^2) \\ \gcd(D, W)=1}} \frac{\overline{f_i}(D)}{D} \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|f_i|}(T; \widetilde{W}q, A'). \quad (8.4)$$

Since

$$\begin{aligned}
& \sum_{\substack{D \leq \exp((\log \log T)^2) \\ \gcd(D, W)=1}} \frac{\overline{f_i}(D)}{D} \leq \prod_{j \neq i} \prod_{w(N) < p \leq \exp((\log \log T)^2)} \left( 1 + \frac{|f_j(p)|}{p} + \frac{|f_j(p^2)|}{p^2} + \dots \right) \\
&\ll \exp \left( \sum_{p \leq \exp((\log \log T)^2)} \frac{|f(p)|}{p} \right) \ll (\log \log T)^{2H},
\end{aligned}$$

this completes the proof of the second part of the lemma.  $\square$

**8.3. Montgomery–Vaughan approach.** From now on we assume that  $D$  is unexceptional, that is,  $D \sim 2^k$  for  $k \geq (\log \log T)^2 / \log 2$  and  $D \notin \mathcal{B}_{2^k}$ , where  $\mathcal{B}_{2^k}$  is the exceptional set from Section 8.1. To bound the inner sum of (8.3) for unexceptional  $D$ , we shall employ the strategy Montgomery and Vaughan followed in their improvement [22] of a result of Daboussi, and begin by introducing a factor  $\log n$  into the average. This will later allow us to reduce matters to understanding equidistribution along sequences defined in terms of primes. We set  $h = f_i$ . Cauchy–Schwarz and several integral comparisons show that

$$\begin{aligned} & \sum_{n \leq T/(D\widetilde{W}q)} \mathbf{1}_I(Dn + D'') h(\widetilde{W}qn + D') F(g(Dn + D'')\Gamma) \log \left( \frac{(T/D)}{\widetilde{W}qn + D'} \right) \\ & \leq \left( \sum_{n \leq T/(D\widetilde{W}q)} \left( \log(T/(D\widetilde{W}q)) - \log n \right)^2 \right)^{1/2} \left( \sum_{n \leq T/(D\widetilde{W}q)} h^2(\widetilde{W}qn + D') \right)^{1/2} \\ & \ll \frac{T}{D\widetilde{W}q} \sqrt{\frac{D\widetilde{W}q}{T} \sum_{n \leq T/(D\widetilde{W}q)} h^2(\widetilde{W}qn + D')}, \end{aligned}$$

and hence, invoking  $D \leq T^{1-1/t}$ ,

$$\begin{aligned} & \sum_{\substack{n \leq T/(D\widetilde{W}q) \\ Dn + D'' \in I}} h(\widetilde{W}qn + D') F(g(Dn + D'')\Gamma) \\ & \ll_t \frac{1}{\log T} \sqrt{\frac{D\widetilde{W}q}{T} \sum_{n \leq \frac{T}{D\widetilde{W}q}} h^2(\widetilde{W}qn + D')} \\ & \quad + \frac{1}{\log T} \left| \frac{D\widetilde{W}q}{T} \sum_{\substack{n \leq T/(D\widetilde{W}q) \\ Dn + D'' \in I}} h(\widetilde{W}qn + D') F(g(Dn + D'')\Gamma) \log(\widetilde{W}qn + D') \right|. \end{aligned} \tag{8.5}$$

By Definition 1.2, condition (2), the contribution of the first term in the bound above to (8.3) is at most

$$O_t \left( (\log T)^{-\theta_f} S_{|f|}(T; \widetilde{W}q, A + \widetilde{W}r) \right)$$

Note that this error term is negligible in view of the bound stated in Proposition 5.3.

It remains to estimate the second term in the bound from (8.5). It will be convenient to abbreviate

$$g_D(n) := g(Dn + D''),$$

and to introduce the discrete interval  $I_D$ , defined via  $n \in I_D$  if and only if  $Dn + D'' \in I$ . Furthermore, let  $P_D$  denote the progression for which  $n \in P_D$  if and only if  $\frac{n - D'}{\widetilde{W}q} \in I_D$ .

Since  $\log n = \sum_{m|n} \Lambda(m)$ , our task is to bound

$$\left| \frac{D\widetilde{W}q}{T \log T} \sum_{\substack{mn \leq T/D \\ mn \equiv D' \pmod{\widetilde{W}q}}} \mathbf{1}_{P_D}(nm) h(nm) \Lambda(m) F\left(g_D\left(\frac{nm - D'}{\widetilde{W}q}\right) \Gamma\right) \right|. \quad (8.6)$$

We may, in fact, restrict the summation in (8.6) to pairs  $(m, n)$  of co-prime integers for which  $m = p$  is prime. To see this, recall that  $F$  is 1-bounded and observe that

$$\begin{aligned} \sum_{\substack{nm \leq T/D \\ \Omega(m) \geq 2 \text{ or } \gcd(n, m) > 1 \\ mn \equiv D' \pmod{\widetilde{W}q}}} |h(nm)| \Lambda(m) &\leq 2 \sum_p \sum_{k \geq 2} k \log p \sum_{\substack{n \leq T/D, p^k \| n \\ n \equiv D' \pmod{\widetilde{W}q}}} |h(n)| \\ &\leq 2 \sum_{p > w(N)} \sum_{k \geq 2} H^k k \log p \sum_{\substack{n \leq T/(Dp^k) \\ p^k n \equiv D' \pmod{\widetilde{W}q}}} |h(n)|. \end{aligned}$$

By Definition 1.2 part (3), the order of the mean value of  $h$  is stable under small perturbations, i.e. if  $p^k \leq (T/D)^{1/2}$ , then

$$\sum_{\substack{n \leq T/(Dp^k) \\ p^k n \equiv D' \pmod{\widetilde{W}q}}} |h(n)| \ll \frac{1}{p^k} \max_{A' \in (\mathbb{Z}/q\widetilde{W}\mathbb{Z})^*} \sum_{\substack{n \leq T/D \\ n \equiv A' \pmod{\widetilde{W}q}}} |h(n)|.$$

The contribution of such  $p^k$  is bounded as follows. Provided  $\varepsilon > 0$  is sufficiently small (e.g.  $\varepsilon = \frac{1}{4}$ ), and  $N$ , and hence  $w(N)$  is sufficiently large, we have

$$\begin{aligned} &\ll \sum_{p > w(N)} \sum_{\substack{k \geq 2 \\ p^k \leq (T/D)^{1/2}}} \frac{H^k \log p^k}{p^k} \max_{A' \in (\mathbb{Z}/q\widetilde{W}\mathbb{Z})^*} \sum_{\substack{n \leq T/D \\ n \equiv A' \pmod{\widetilde{W}q}}} |h(n)| \\ &\ll \sum_{p > w(N)} \frac{1}{p^{2-\varepsilon}} \max_{A' \in (\mathbb{Z}/q\widetilde{W}\mathbb{Z})^*} \sum_{\substack{n \leq T/D \\ n \equiv A' \pmod{\widetilde{W}q}}} |h(n)| \\ &\ll \frac{1}{w(N)^{1-\varepsilon}} \max_{A' \in (\mathbb{Z}/q\widetilde{W}\mathbb{Z})^*} \sum_{\substack{n \leq T/D \\ n \equiv A' \pmod{\widetilde{W}q}}} |h(n)| \\ &\ll \frac{T}{D\widetilde{W}q} \frac{1}{w(N)^{1-\varepsilon}} \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|h|}(T/D; \widetilde{W}q, A'). \end{aligned}$$

The remaining sum over  $p^k > (T/D)^{1/2}$  satisfies:

$$\begin{aligned}
& \sum_{p > w(N)} \sum_{\substack{k \geq 2 \\ p^k > (T/D)^{1/2}}} H^k \log p^k \sum_{\substack{n \leq T/(Dp^k) \\ p^k n \equiv D' \pmod{\widetilde{W}q}}} |h(n)| \\
& \ll \sum_{p > w(N)} \sum_{\substack{k \geq 2 \\ p^k > (T/D)^{1/2}}} H^k \log p^k \frac{T}{Dp^k} \left( \log \frac{T}{Dp^k} \right)^{O(H)} \\
& \ll \frac{T}{D} \left( \log \frac{T}{D} \right)^{O(H)} \sum_{p > w(N)} \sum_{\substack{k \geq 2 \\ p^k > (T/D)^{1/2}}} \frac{H^k \log p^k}{p^k} \\
& \ll \frac{T}{D} \left( \log \frac{T}{D} \right)^{O(H)} \sum_{p > w(N)} \sum_{\substack{k \geq 2 \\ p^k > (T/D)^{1/2}}} p^{-k+\varepsilon} \\
& \ll \frac{T}{D} \left( \log \frac{T}{D} \right)^{O(H)} \sum_{p > w(N)} (T/D)^{-\varepsilon} p^{-2+2\varepsilon} \\
& \ll \left( \frac{T}{D} \right)^{1-\varepsilon} \left( \log \frac{T}{D} \right)^{O(H)},
\end{aligned}$$

again, provided  $\varepsilon > 0$  is sufficiently small, e.g.  $\varepsilon = \frac{1}{4}$ , and  $N$ , and hence  $w(N)$ , sufficiently large.

Thus, the total contribution to (8.3) of pairs  $(m, n)$  that are not of the form  $(m, p)$ , where  $p$  is prime and does not divide  $m$ , is bounded by

$$\begin{aligned}
& \frac{T}{\widetilde{W}q} \sum_{i=1}^t \sum_k \sum_{D \sim 2^k} \sum_{d_1 \dots \hat{d}_i \dots d_t = D} \left( \prod_{j \neq i} |f_j(d_j)| \right) \frac{1}{\log T} \frac{1}{w(N)^{1-\varepsilon}} \frac{1}{D} \max_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} S_{|h|}(T/D; \widetilde{W}, A') \\
& \leq \frac{T}{\widetilde{W}q} \frac{1}{\log T} \frac{1}{w(N)^{1-\varepsilon}} \max_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} S_{|f_1| * \dots * |f_t|}(T; \widetilde{W}, A'),
\end{aligned}$$

which is negligible in view of the bound claimed in Proposition 5.3.

It therefore suffices to bound the expression

$$\frac{D\widetilde{W}q}{T} \sum_{\substack{mp \leq T/D \\ mp \equiv D' \pmod{\widetilde{W}q}}} \mathbf{1}_{P_D}(mp) h(m) h(p) \Lambda(p) F\left(g\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right).$$

We shall split this summation into large and small divisors with respect to the parameter

$$X = X(D) = \left( \frac{T}{D} \right)^{1-1/(\log \frac{T}{D})^{\frac{U-1}{U}}},$$



where  $1 \leq U \ll 1$  is any integer satisfying  $U^{-1} < \theta_f/2$ , e.g.

$$U = \max\{1, \lceil 2/\theta_f \rceil\}.$$

With this choice of  $X$  we obtain

$$\begin{aligned} & \frac{\widetilde{W}qD}{T} \sum_{\substack{m < X \\ \gcd(m, \widetilde{W})=1}} \sum_{\substack{p \leq T/(mD) \\ p \equiv D\overline{m} \pmod{\widetilde{W}}}} \mathbf{1}_{P_D}(mp)h(m)h(p)\Lambda(p)F\left(g\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \\ & + \frac{\widetilde{W}qD}{T} \sum_{\substack{m > X \\ \gcd(m, \widetilde{W})=1}} \sum_{\substack{p \leq T/(mD) \\ p \equiv D\overline{m} \pmod{\widetilde{W}}}} \mathbf{1}_{P_D}(mp)h(m)h(p)\Lambda(p)F\left(g\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right). \end{aligned} \quad (8.7)$$

In order to analyse these expressions, we dyadically decompose in each of the two terms the sum with shorter summation range. The cut-off parameter  $X$  is chosen in such a way that one of the dyadic decompositions is of short length, depending on  $U$ . Indeed, we have

$$\log_2 X \sim \log_2(T/D),$$

while

$$\log_2 \frac{T}{DX} = \frac{(\log \frac{T}{D})^{1/U}}{\log 2}.$$

Let

$$T_0 = \exp((\log \log T)^2).$$

Then the two sums from (8.7) decompose as

$$\begin{aligned} & \frac{\widetilde{W}qD}{T} \sum_{\substack{m < T_0 \\ \gcd(m, \widetilde{W})=1}} \sum_{\substack{p \leq T/(mD) \\ p \equiv D\overline{m} \pmod{\widetilde{W}}}} \mathbf{1}_{P_D}(mp)h(m)h(p)\Lambda(p)F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \\ & + \frac{\widetilde{W}qD}{T} \sum_{j=1}^{\log_2(X/T_0)} \sum_{\substack{m \sim 2^{-j}X \\ \gcd(m, \widetilde{W})=1}} \sum_{\substack{p \leq T/(mD) \\ p \equiv D\overline{m} \pmod{\widetilde{W}}}} \mathbf{1}_{P_D}(mp)h(m)h(p)\Lambda(p)F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \end{aligned} \quad (8.8)$$

and

$$\begin{aligned} & \frac{\widetilde{W}qD}{T} \left\{ \sum_{\substack{m > X \\ \gcd(m, \widetilde{W})=1}} \sum_{\substack{p \leq \min(T/(mD), T_0) \\ p \equiv D\overline{m} \pmod{\widetilde{W}}}} \mathbf{1}_{P_D}(mp)h(m)h(p)\Lambda(p)F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \right. \\ & \left. + \sum_{j=1}^{\log_2(T/(XD T_0))} \sum_{\substack{m > X \\ \gcd(m, \widetilde{W})=1}} \sum_{\substack{p \sim 2^{-j}T/(XD) \\ p \equiv D\overline{m} \pmod{\widetilde{W}}}} \mathbf{1}_{pm < T/D} \mathbf{1}_{P_D}(mp)h(m)h(p)\Lambda(p)F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \right\}. \end{aligned} \quad (8.9)$$

We now analyse the contribution from these four sums to (8.3) in turn, beginning with the two short sums up to  $T_0$ .

**8.4. Short sums.** The following lemma gives bounds on the contribution of the short sums in (8.8) and (8.9) to (8.3).

**Lemma 8.2.** *Writing  $\bar{f}_i(n) = |f_1 * \cdots * \widehat{f_i} * \cdots * f_t(n)|$ , we have*

$$\begin{aligned} & \sum_{D \leq T^{1-1/t}} \left| \frac{\bar{f}_i(D)}{\log T} \right| \sum_{\substack{m < T_0 \\ \gcd(m, W)=1}} \sum_{\substack{p \leq T/(mD) \\ p \equiv D\bar{m} \pmod{\widetilde{W}q}}} \mathbf{1}_{P_D}(mp) h(m) h(p) \Lambda(p) F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right) \Gamma\right) \\ & \ll \frac{T}{\widetilde{W}q} (\log \log T)^2 \frac{Wq}{\phi(Wq)} \frac{1}{\log T} \exp\left(\sum_{w(N) < p < T} \frac{|f_1(p) + \cdots + f_t(p) - f_i(p)|}{p}\right) \end{aligned}$$

and

$$\begin{aligned} & \sum_{D \leq T^{1-1/t}} \left| \frac{\bar{f}_i(D)}{\log T} \right| \sum_{\substack{m > X \\ \gcd(m, W)=1}} \sum_{\substack{p \leq \min(\frac{T}{Dm}, T_0) \\ p \equiv D\bar{m} \pmod{\widetilde{W}q}}} \mathbf{1}_{P_D}(mp) h(m) h(p) \Lambda(p) F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right) \Gamma\right) \\ & \ll \frac{T}{\widetilde{W}q} \frac{(\log \log T)^2}{\log T} \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|f_1| \cdots |f_t|}(T; \widetilde{W}, A'). \end{aligned}$$

**Remark.** Note that both the above bounds make sense in the case where  $t = 1$ .

*Proof.* The short sum in (8.8) satisfies

$$\begin{aligned} & \frac{1}{\log T} \left| \frac{\widetilde{W}qD}{T} \sum_{\substack{m < T_0 \\ \gcd(m, W)=1}} \sum_{\substack{p \leq T/(mD) \\ p \equiv D\bar{m} \pmod{\widetilde{W}q}}} \mathbf{1}_{P_D}(mp) h(m) h(p) \Lambda(p) F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right) \Gamma\right) \right| \\ & \ll \frac{1}{\log T} \frac{\widetilde{W}qD}{T} \sum_{\substack{m < T_0 \\ \gcd(m, W)=1}} \sum_{\substack{p \leq T/(mD) \\ p \equiv D\bar{m} \pmod{\widetilde{W}q}}} |\mathbf{1}_{P_D}(mp) h(m) h(p) \Lambda(p)| \\ & \ll \frac{\widetilde{W}q}{\phi(\widetilde{W}q)} \frac{1}{\log T} \sum_{\substack{m < T_0 \\ \gcd(m, W)=1}} \frac{|h(m)|}{m} \ll \frac{Wq}{\phi(Wq)} \frac{1}{\log T} \exp\left(\sum_{w(N) < p < T_0} \frac{|h(p)|}{p}\right) \\ & \ll (\log \log T)^2 \frac{Wq}{\phi(Wq)} \frac{1}{\log T}, \end{aligned}$$

where we made use of part (3) of Definition 1.2 This yields the first bound, since

$$\sum_{D \leq T^{1-1/t}} \frac{\bar{f}_i(D)}{D} \ll \exp\left(\sum_{w(N) < p < T} \frac{|f_1(p) + \cdots + f_t(p) - f_i(p)|}{p}\right),$$

which continues to hold when  $t = 1$ .

The short sum in (8.9) is bounded by

$$\begin{aligned}
& \frac{1}{\log T} \left| \frac{\widetilde{W}qD}{T} \sum_{\substack{m > X \\ \gcd(m, W)=1}} \sum_{\substack{p \leq \min(T/(mD), T_0) \\ p \equiv D'\overline{m} \pmod{\widetilde{W}q}}} \mathbf{1}_{P_D}(mp) h(m) h(p) \Lambda(p) F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right) \Gamma\right) \right| \\
& \ll \frac{1}{\log T} \sum_{w(N) < p < T_0} \frac{\Lambda(p)}{p} \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|h|}(T/(pD); \widetilde{W}q, A') \\
& \ll \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|h|}(T/D; \widetilde{W}q, A') \frac{1}{\log T} \sum_{w(N) < p < T_0} \frac{\Lambda(p)}{p} \\
& \ll \frac{\log T_0}{\log T} \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|h|}(T/D; \widetilde{W}q, A') = \frac{(\log \log T)^2}{\log T} \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|h|}(T/D; \widetilde{W}q, A').
\end{aligned}$$

This proves the second part of the lemma, since

$$\sum_{D \leq T^{1-1/t}} \frac{\bar{f}_i(D)}{D} \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|h|}(T/D; \widetilde{W}q, A') \ll \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|f_1| * \dots * |f_t|}(T; \widetilde{W}q, A').$$

□

Since the statement of Proposition 5.3 takes care of both bounds from the above lemma, it remains to bound the contribution from the dyadic parts of (8.8) and (8.9).

**8.5. Large primes.** In this subsection we prove the following lemma which handles the contribution of the dyadic parts of (8.8) to (8.3).

**Lemma 8.3** (Contribution from large primes). *Under the assumptions of Proposition 5.3, the following holds. Let  $E_h^\sharp(T; D, q, j, I)$  denote the expression*

$$\left| \frac{D\widetilde{W}q}{T} \sum_{\substack{m \sim 2^{-j}X \\ \gcd(m, W)=1}} \sum_{\substack{p \leq T/(mD) \\ p \equiv D'\overline{m} \pmod{\widetilde{W}q}}} \mathbf{1}_{P_D}(mp) h(m) h(p) \Lambda(p) F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right) \Gamma\right) \right|.$$

Then, provided the parameter  $E_0$  from Proposition 5.3 is sufficiently large depending on  $d$ ,  $m_G$  and  $H$ , we have

$$\begin{aligned}
& \sum_{i=1}^t \sum_{k=\frac{(\log \log T)^2}{\log 2}}^{(1-1/t)\frac{\log T}{\log 2}} \sum_{D \sim 2^k} \mathbf{1}_{D \notin \mathcal{B}_{2^k}} \sum_{\substack{d_1, \dots, \widehat{d_i}, \dots, d_t \\ D_i = D}} \left( \prod_{i' \neq i} \frac{|f_{i'}(d_{i'})|}{d_{i'}} \right)^{\log_2(X/T_0)} \sum_{j=0}^{\log_2(X/T_0)} \frac{E_{f_i}^\sharp(T; D, q, j, I)}{\log T} \\
& \ll_{d, s, m_G, E, H, U} \left( (\log \log T)^{-1/(2^{s+2} \dim G)} + \frac{Q^{10^s \dim G}}{(\log \log T)^{1/2}} \right) \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} S_{|f_1| * \dots * |f_t|}(T; \widetilde{W}q, A').
\end{aligned} \tag{8.10}$$

**Remark.** Note that this contribution does not exceed the bound (5.2).

The remainder of this subsection is concerned with the proof of (8.10). Considering  $E_h^\sharp(T; D, q, j, I)$  for a fixed value of  $j$ ,  $1 \leq j \leq \log_2 \frac{X}{T_0}$ , the Cauchy–Schwarz inequality yields

$$\begin{aligned}
& \sum_{\substack{m \sim 2^{-j}X \\ \gcd(m, W)=1}} \sum_{\substack{p \leq 2^j T/(XD) \\ p \equiv D\overline{m} \pmod{\widetilde{W}q} \\ mp \in P_D}} \mathbf{1}_{mp \leq N} h(m) h(p) \Lambda(p) F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \\
& \leq \left( \sum_{p \leq 2^j T/(XD)} |h(p)|^2 \Lambda(p) \right)^{1/2} \left( \frac{\widetilde{W}q}{\phi(\widetilde{W}q)} \sum_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} \sum_{\substack{m, m' \sim 2^{-j}X \\ m \equiv m' \equiv A' \pmod{\widetilde{W}q} \\ \gcd(W, m)=1}} h(m) h(m') \right. \\
& \quad \left. \frac{\phi(\widetilde{W}q)}{\widetilde{W}q} \sum_{\substack{p \leq T/(D \max(m, m')) \\ pA' \equiv D' \pmod{\widetilde{W}q} \\ pm, pm' \in P_D}} \Lambda(p) F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \overline{F\left(g_D\left(\frac{pm' - D'}{\widetilde{W}q}\right)\Gamma\right)} \right)^{1/2}.
\end{aligned} \tag{8.11}$$

The first factor is easily seen to equal  $O(1)$ , since  $h(p) \ll_H 1$  at primes. To estimate the second factor, we seek to employ the orthogonality of the ‘ $W$ -tricked von Mangoldt function’ with nilsequences in combination with the fact that Proposition 7.1 implies that for most pairs  $(m, m')$  the product nilsequence that appears in the above expression is equidistributed. For this purpose, we effect the change of variables  $p = \widetilde{W}qn + D'_m$  in the inner sum of the second factor, where  $D'_m$  is such that  $D'_m \equiv D'\overline{m} \pmod{\widetilde{W}q}$ . This yields

$$\begin{aligned}
& \frac{\phi(\widetilde{W}q)}{\widetilde{W}q} \sum_{\substack{p \leq T/(D \max(m, m')) \\ pA' \equiv D' \pmod{\widetilde{W}q} \\ pm, pm' \in P_D}} \Lambda(p) F\left(g_D\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \overline{F\left(g_D\left(\frac{pm' - D'}{\widetilde{W}q}\right)\Gamma\right)} \\
& = \sum_{\substack{n \leq T/(\widetilde{W}qD \max(m, m')) \\ nm + \widetilde{D}_m, nm' + \widetilde{D}_{m'} \in I_D}} \frac{\phi(\widetilde{W}q)}{\widetilde{W}q} \Lambda(\widetilde{W}qn + D'_m) F(g_D(nm + \widetilde{D}_m)\Gamma) \overline{F(g_D(nm' + \widetilde{D}_{m'})\Gamma)},
\end{aligned} \tag{8.12}$$

for suitable values of  $0 \leq \widetilde{D}_m < m$ ,  $0 \leq \widetilde{D}_{m'} < m'$ .

Recall that for any fixed unexceptional value of  $D$ , the finite sequence

$$(g_D(n)\Gamma)_{n \leq T/(Dq)}$$

is totally  $\delta(N)^{c_1 E_0}$ -equidistributed. Thus, applying Proposition 7.1 with  $g = g_D$  and with  $E_2 = c_1 E_0$ , we obtain for every integer

$$K \in [T_0, X]$$

an exceptional set  $\mathcal{E}_K$  of size  $\#\mathcal{E}_K \ll \delta(T)^{O(c_1 c_2 E_0)} K^2$  such that for all pairs  $m, m' \sim K$  with  $m \equiv m' \pmod{W}$  and  $(m, m') \notin \mathcal{E}_K$  the following estimate holds:

$$\left| \sum_{\substack{n \leq T/(\widetilde{W}qD \max(m, m')) \\ nm + \widetilde{D}_m, nm' + \widetilde{D}_{m'} \in I_D}} F(g_D(nm + \widetilde{D}_m)\Gamma) \overline{F(g_D(nm' + \widetilde{D}_{m'})\Gamma)} \right| < \frac{(1 + \|F\|_{\text{Lip}})\delta(N)^{c_1 c_2 E_0} T}{KqD}.$$

Before we continue with the analysis of (8.12), we prove a quick lemma that will allow us to handle the contribution of exceptional sets  $\mathcal{E}_K$  in the proof of Lemma 8.3.

**Lemma 8.4.** *Suppose  $j \leq \log_2(X/T_0)$  and let  $\mathcal{E}_{D,K}$  denote the exceptional set from Proposition 7.1 with  $g = g_D$ . Then, provided  $E_0$  is sufficiently large, we have*

$$\begin{aligned} & \frac{1}{\phi(\widetilde{W}q)} \sum_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} \sum_{\substack{m, m' \sim 2^{-j}X \\ m \equiv m' \pmod{\widetilde{W}q}}} |h(m)h(m')| \mathbf{1}_{(m, m') \in \mathcal{E}_{D, 2^{-j}X}} \\ & \ll \delta(N)^{O(c_1 c_2 E_0)} \left( \max_{A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*} \sum_{\substack{m \sim 2^{-j}X \\ m \equiv A' \pmod{\widetilde{W}q}}} |h(m)| \right)^2. \end{aligned}$$

*Proof.* Cauchy–Schwarz and an application of part (1) of Definition 1.2 yield for each residue  $A' \in (\mathbb{Z}/\widetilde{W}q\mathbb{Z})^*$  that

$$\begin{aligned} & \sum_{\substack{m, m' \sim 2^{-j}X \\ m \equiv m' \pmod{\widetilde{W}q}}} |h(m)h(m')| \mathbf{1}_{(m, m') \in \mathcal{E}_{D, 2^{-j}X}} \\ & \ll 2^{-j}X \delta(N)^{O(c_1 c_2 E_0)} \left( \sum_{\substack{m, m' \sim 2^{-j}X \\ m \equiv m' \pmod{\widetilde{W}q}}} |h(m)|^2 |h(m')|^2 \right)^{1/2} \\ & \ll 2^{-j}X \delta(N)^{O(c_1 c_2 E_0)} \sum_{\substack{m \sim 2^{-j}X \\ m \equiv A' \pmod{\widetilde{W}q}}} |h(m)|^2 \\ & \ll \delta(N)^{O(c_1 c_2 E_0)} (\log 2^{-j}X)^2 \left( \sum_{\substack{m \sim 2^{-j}X \\ m \equiv A' \pmod{\widetilde{W}q}}} |h(m)| \right)^2. \end{aligned}$$

Since  $\delta(N) \leq (\log N)^{-1}$ , any sufficiently large choice of  $E_0$  guarantees that

$$\delta(N)^{O(c_1 c_2 E_0)} (\log 2^{-j}X)^2 \leq \delta(N)^{O(c_1 c_2 E_0)}$$

holds. This completes the proof.  $\square$

As a final tool for the proof of Lemma 8.3, we require an explicit bound on the correlation of the ‘ $W$ -tricked von Mangoldt function’ with nilsequences. The following lemma, which will be proved in Appendix A building on the proof of [12, Proposition 10.2], provides such bounds in our specific setting.

**Lemma 8.5.** *Let  $G/\Gamma$  be an  $s$ -step nilmanifold, let  $G_\bullet$  be a filtration of  $G$  and let  $\mathcal{X}$  be a  $Q$ -rational Mal'cev basis adapted to it. Let  $\Lambda' : \mathbb{N} \rightarrow \mathbb{R}$  be the restriction of the ordinary von Mangoldt function to primes, that is,  $\Lambda'(p^k) = 0$  whenever  $k > 1$ . Let  $W = W(x)$ , let  $q'$  and  $b'$  be integers such that  $0 < b' < Wq' \leq (\log x)^E$  and  $\gcd(W, b') = 1$  hold. Let  $\alpha \in (0, 1)$ . Then, for every  $y \in [\exp((\log x)^\alpha), x]$  and for every polynomial sequence  $g \in \text{poly}(\mathbb{Z}, G_\bullet)$ , the following estimate holds:*

$$\left| \sum_{n \leq y} \frac{\phi(Wq')}{Wq'} \Lambda'(Wq'n + b') F(g(n)\Gamma) \right| \ll_{\alpha, s, E, \|F\|_{\text{Lip}}} \left| \sum_{n \leq y} F(g(n)\Gamma) \right| + y \mathcal{E}(x),$$

where

$$\mathcal{E}(x) := (\log \log x)^{-1/(2^{s+2} \dim G)} + \frac{Q^{10^s \dim G}}{(\log \log x)^{1/2}}.$$

Employing Lemma 8.5 for the upper endpoint of an interval  $[y_0, y_1]$ , and either a trivial estimate or the lemma for the lower endpoint, say, depending on whether or not  $y_0 \leq y_1^{1/2}$ , we obtain as immediate consequence that

$$\begin{aligned} & \left| \sum_{y_0 \leq n \leq y_1} \frac{\phi(Wq')}{Wq'} \Lambda'(Wq'n + b') F(g(n)\Gamma) \right| \\ & \ll_{\alpha, s, E, \|F\|_{\text{Lip}}} \left| \sum_{n \leq y_0} F(g(n)\Gamma) \right| + \left| \sum_{n \leq y_1} F(g(n)\Gamma) \right| + y_1^{1/2} + y_1 \mathcal{E}(x). \end{aligned} \quad (8.13)$$

for any  $0 < y_0 < y_1 \leq x$  such that  $y_1 \geq \exp((\log x)^\alpha)$ .

To continue with the proof of Lemma 8.3, we now return to the right hand side of (8.12). Recall that  $n \in I_D$  if and only if  $Dn + D'' \in I$ . Since  $I$  is a discrete interval,  $I_D$  is a discrete interval too and, for  $m, m' \sim 2^{-j}X$ , we have

$$\#\{n \in \mathbb{N} : nm + \tilde{D}_m \in I_D\} \ll |I_D| 2^j / X \ll |I| 2^j / (DX) \leq T 2^j / (DX \widetilde{W}q)$$

and, similarly,  $\#\{n \in \mathbb{N} : nm' + \tilde{D}_{m'} \in I_D\} \ll |I_D| 2^j / X \ll T 2^j / (DX \widetilde{W}q)$ . We now consider separately the cases of  $(m, m') \notin \mathcal{E}_{D, 2^{-j}X}$  for which

$$I_{m, m'} = \left\{ n \in \mathbb{N} : \begin{array}{l} nm + \tilde{D}_m \in I_D, \\ nm' + \tilde{D}_{m'} \in I_D \end{array} \right\} \leq \delta(N) 2^j T / (DX \widetilde{W}q) \quad (8.14)$$

holds and those of  $(m, m') \notin \mathcal{E}_{D, 2^{-j}X}$  for which  $I_{m, m'} > \delta(N) 2^j T / (DX \widetilde{W}q)$ . In the former case, the trivial bound asserts

$$\sum_{\substack{n \leq \frac{T}{\widetilde{W}q D \max(m, m')} \\ nm + \tilde{D}_m \in I \\ nm' + \tilde{D}_{m'} \in I}} \frac{\phi(\widetilde{W}q)}{\widetilde{W}q} \Lambda(\widetilde{W}qn + D'_m) F(g_D(nm + \tilde{D}_m)\Gamma) \overline{F(g_D(nm' + \tilde{D}_{m'})\Gamma)} \leq \frac{\delta(N) T 2^j}{DX \widetilde{W}q}.$$

In the latter case, we seek to apply (8.13) with  $[y_0, y_1] = I_{m,m'}$  and  $x = N = T^{1+o(1)}$ . Note that

$$\begin{aligned}
\log y_1 &\geq \log \left( \delta(N) T 2^j / (DX \widetilde{W} q) \right) \\
&\geq \log \frac{T}{DX} + j \log 2 + E_0 \log \delta(N) - \log(\widetilde{W} q) \\
&\geq (\log(T/D))^{1/U} + j \log 2 - O_{d, \dim G, E, E_0, U}(1) \log \log N - 2E \log \log T \\
&\geq \left( \frac{1}{t} \log T \right)^{1/U} - O_{d, \dim G, E, E_0, U}(1) \log \log N - 2E \log \log T \\
&\gg_{d, \dim G, E, E_0, H, U} (\log T)^{1/U},
\end{aligned}$$

where we used the definition of  $X$ , the assumptions that Proposition 5.3 makes on  $\delta$ , and the fact that  $t \leq H$ . Thus, the conditions of Lemma 8.5 are satisfied for every sufficiently large  $T$  (with respect to  $d, \dim G, E, E_0, H$  and  $U$ ) when choosing  $\alpha = 1/(2U)$ . Hence, (8.13) yields the following estimate for the interval  $[y_0, y_1] = I_{m,m'}$ :

$$\begin{aligned}
&\sum_{\substack{n \leq T/(\widetilde{W} q D \max(m, m')) \\ nm + \widetilde{D}_m, nm' + \widetilde{D}_{m'} \in I}} \frac{\phi(\widetilde{W} q)}{\widetilde{W} q} \Lambda(\widetilde{W} q n + D'_m) F(g_D(nm + \widetilde{D}_m) \Gamma) \overline{F(g_D(nm' + \widetilde{D}_{m'}) \Gamma)} \\
&\ll_{d, \dim G, s, E, E_0, H, U} \sum_{n \leq y_0} F(g_D(nm + \widetilde{D}_m) \Gamma) \overline{F(g_D(nm' + \widetilde{D}_{m'}) \Gamma)} \\
&\quad + \sum_{n \leq y_1} F(g_D(nm + \widetilde{D}_m) \Gamma) \overline{F(g_D(nm' + \widetilde{D}_{m'}) \Gamma)} + \frac{T 2^j}{DX \widetilde{W} q} \mathcal{E}(T).
\end{aligned}$$

Proposition 7.1 shows that the right hand side is small for most pairs  $(m, m')$ . Indeed, together with Proposition 7.1, the above implies that (8.11) is bounded above by

$$\begin{aligned}
&\ll_{d, \dim G, s, E, E_0, H, U} \left( \frac{T 2^j}{DX} \right)^{1/2} \left( \frac{\widetilde{W} q}{\phi(\widetilde{W} q)} \sum_{A' \in (\mathbb{Z}/\widetilde{W} q \mathbb{Z})^*} \sum_{\substack{m, m' \sim 2^{-j} X \\ m \equiv m' \equiv A' \pmod{\widetilde{W} q}}} |h(m) h(m')| \times \right. \\
&\quad \left. \times \frac{T 2^j}{\widetilde{W} q DX} \left( \delta(N)^{O(c_1 c_2 E_0)} + \mathbf{1}_{(m, m') \in \mathcal{E}_{D, 2^{-j} X}} + \mathcal{E}(T) \right) \right)^{1/2},
\end{aligned}$$

An application of Lemma 8.4 show that this in turn is bounded by

$$\begin{aligned} & \ll_{d, \dim G, s, E, E_0, H, U} \frac{T}{\widetilde{W}_q D} \left( \max_{A' \in (\mathbb{Z}/\widetilde{W}_q \mathbb{Z})^*} \left( \frac{\widetilde{W}_q 2^j}{X} \sum_{\substack{m \sim 2^{-j} X \\ m \equiv A' \pmod{\widetilde{W}_q}}} |h(m)| \right)^2 \right. \\ & \quad \left. \times \left( \delta(N)^{O(c_1 c_2 E_0)} + \mathcal{E}(T) \right) \right)^{1/2} \\ & \ll_{d, \dim G, s, E, E_0, H, U} \frac{T}{\widetilde{W}_q D} \left( \delta(N)^{O(c_1 c_2 E_0)} + \mathcal{E}(T) \right) \max_{A' \in (\mathbb{Z}/\widetilde{W}_q \mathbb{Z})^*} S_{|h|}(2^{-j} X; \widetilde{W}_q, A'). \end{aligned}$$

Summing the above expression over  $j \leq \log_2(\frac{X}{T_0})$  and taking into account the factor  $(\log T)^{-1}$ , we deduce that the inner sum in (8.10) is bounded by

$$\ll_{d, \dim G, s, E, E_0, H, U} \left( \delta(N)^{O(c_1 c_2 E_0)} + \mathcal{E}(T) \right) \frac{1}{\log T} \sum_{j=0}^{\log_2(X/T_0)} \max_{A' \in (\mathbb{Z}/\widetilde{W}_q \mathbb{Z})^*} S_{|h|}(2^{-j} X; \widetilde{W}_q, A').$$

Properties (3) and (4) of Definition 1.2 show that this in turn is at most

$$\begin{aligned} & \ll_{d, \dim G, s, E, E_0, H, U} \left( \delta(N)^{O(c_1 c_2 E_0)} + \mathcal{E}(T) \right) \max_{A' \in (\mathbb{Z}/\widetilde{W}_q \mathbb{Z})^*} S_{|h|}(X; \widetilde{W}_q, A') \\ & \ll_{d, \dim G, s, E, E_0, H, U} \left( \delta(N)^{O(c_1 c_2 E_0)} + \mathcal{E}(T) \right) \max_{A' \in (\mathbb{Z}/\widetilde{W}_q \mathbb{Z})^*} S_{|h|}(T/D; \widetilde{W}_q, A'). \end{aligned}$$

Since  $\delta(N) \leq (\log N)^{-1}$ , it suffices to choose  $E_0$  sufficiently large in terms of  $d$  and  $m_0$  in order to obtain

$$\delta(N)^{O(c_1 c_2 E_0)} + \mathcal{E}(T) \ll (\log N)^{-1} + \mathcal{E}(T) \ll \mathcal{E}(T).$$

This completes the proof of Lemma 8.3.

**8.6. Small primes.** To complete the proof of Proposition 5.3, it remains to bound the contribution of the dyadic parts of (8.9) to (8.3). This is achieved by the following lemma.

**Lemma 8.6** (Contribution from small primes). *Let  $E_h^\flat(T; D, q, j, I)$  denote the expression*

$$\left| \frac{D\widetilde{W}_q}{T} \sum_{\substack{m > X \\ \gcd(m, W)=1}} \sum_{\substack{p \sim 2^{-j} T/(XD) \\ p \equiv D' \overline{m} \pmod{\widetilde{W}_q}}} \mathbf{1}_{pm < T/D} \mathbf{1}_{I_D}(mp) h(m) h(p) \Lambda(p) F\left(g_D\left(\frac{pm - D'}{\widetilde{W}_q}\right) \Gamma\right) \right|.$$

Then

$$\begin{aligned} & \sum_{i=1}^t \sum_{k=\frac{(\log \log T)^2}{\log 2}}^{(1-1/t)\frac{\log T}{\log 2}} \sum_{D \sim 2^k} \mathbf{1}_{D \notin \mathcal{B}_{2^k}} \sum_{\substack{d_1, \dots, \widehat{d_i}, \dots, d_t \\ D_i = D}} \left( \prod_{j \neq i} \frac{|f_j(d_j)|}{d_j} \right)^{\log_2(T/(XD T_0))} \sum_{j=0}^{\log_2(T/(XD T_0))} \frac{E_{f_i}^\flat(T; D, q, j, P')}{\log T} \\ & \ll (\log T)^{-\theta_f/2} \max_{r \in (\mathbb{Z}/\widetilde{W} \mathbb{Z})^*} S_{|f_1| \dots |f_t|}(T; \widetilde{W}; r). \end{aligned}$$



*Proof.* Applying Cauchy–Schwarz to the expression  $E_h^\flat(T; D, q, j, I_D)$  for a fixed value of  $j$ ,  $0 \leq j \leq \log_2(T/(XDT_0))$ , we obtain

$$\begin{aligned}
& \frac{\widetilde{W}D}{T} \sum_{\substack{m > X \\ \gcd(m, W)=1}} \sum_{\substack{p \sim 2^{-j}T/(XD) \\ p \equiv D'\overline{m} \pmod{\widetilde{W}q}}} 1_{pm < T/D} h(m)h(p)\Lambda(p)F\left(g\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \mathbf{1}_{I_D}(mp) \quad (8.15) \\
& \leq \left( \frac{W}{\phi(W)} \frac{1}{2^j X} \sum_{\substack{X < m < 2^j X \\ \gcd(m, W)=1}} |h(m)|^2 \right)^{1/2} \\
& \quad \times \left( \phi(\widetilde{W}q) \left( \frac{2^j XD}{T} \right)^2 \sum_{\substack{p, p' \sim 2^{-j}T/(XD) \\ p \equiv p' \pmod{\widetilde{W}q}}} h(p)h(p')\Lambda(p)\Lambda(p') \right. \\
& \quad \left. \frac{\widetilde{W}q}{X2^j} \sum_{\substack{X < m < T/(D \max(p, p')) \\ mp \equiv D' \pmod{\widetilde{W}q} \\ pm \in I_D}} F\left(g\left(\frac{pm - D'}{\widetilde{W}q}\right)\Gamma\right) \overline{F\left(g\left(\frac{p'm - D'}{\widetilde{W}q}\right)\Gamma\right)} \right)^{1/2}.
\end{aligned}$$

We estimate the second factor trivially as  $O(1)$  by using the bounds  $|h(p)h(p')| \ll 1$  and  $\|F\|_\infty = \|\overline{F}\|_\infty \leq 1$ . Thus, (8.15) is bounded by

$$\begin{aligned}
& \left( \frac{W}{\phi(W)} \frac{1}{2^j X} \sum_{\substack{X < m < 2^j X \\ \gcd(m, W)=1}} |h(m)|^2 \right)^{1/2} = \left( \frac{1}{\phi(\widetilde{W})} \sum_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} \frac{\widetilde{W}}{2^j X} \sum_{\substack{X < m < 2^j X \\ m \equiv A' \pmod{\widetilde{W}}}} |h(m)|^2 \right)^{1/2} \\
& \leq \max_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} \left( \frac{\widetilde{W}}{2^j X} \sum_{\substack{m < 2^j X \\ m \equiv A' \pmod{\widetilde{W}}}} |h(m)|^2 \right)^{1/2}.
\end{aligned}$$

Note that  $X \leq 2^j X \leq T/(DT_0)$ , where

$$X = \left( \frac{T}{D} \right)^{1 - 1/(\log \frac{T}{D})^{(U-1)/U}} \gg (T/D)^{1/2}$$

and

$$\frac{T}{DT_0} = \left( \frac{T}{D} \right)^{1 - (\log \log \frac{T}{D})^2 / (\log \frac{T}{D})}.$$

Thus, the local stability of average values we assumed in property (3) of Definition 1.2 implies that the above is bounded by

$$\ll \max_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} \left( \frac{\widetilde{W}}{2^j X} \sum_{\substack{m < T/D \\ m \equiv A' \pmod{\widetilde{W}}}} |h(m)|^2 \right)^{1/2}.$$

The summation range in  $j$  is short: it is bounded by

$$\log_2(T/(XDT_0)) \ll (\log T)^{1/U} \ll (\log T)^{\theta_f/2}.$$

Thus, invoking property (2) of Definition 1.2, we deduce that

$$\begin{aligned} & \sum_{i=1}^t \sum_{k=1}^{(1-1/t)\frac{\log T}{\log 2}} \sum_{D \sim 2^k} \mathbf{1}_{D \notin \mathcal{B}_{2^k}} \sum_{\substack{d_1, \dots, \widehat{d_i}, \dots, d_t \\ D_i = D}} \left( \prod_{j \neq i} \frac{|f_j(d_j)|}{d_j} \right)^{\log_2(T/(XDT_0))} \sum_{j=0}^{\log_2(T/(XDT_0))} \frac{E_{f_i}^b(T; D, q, j, I)}{\log T} \\ & \ll (\log T)^{-1 + \frac{\theta_f}{2}} \\ & \sum_{i=1}^t \sum_{D \leq T^{1-1/t}} \sum_{\substack{d_1, \dots, \widehat{d_i}, \dots, d_t \\ D_i = D}} \left( \prod_{j \neq i} \frac{|f_j(d_j)|}{d_j} \right) \max_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} \left( \frac{D}{T} \sum_{\substack{m < T/D \\ m \equiv A' \pmod{\widetilde{W}}}} |f_i(m)|^2 \right)^{1/2} \\ & \ll (\log T)^{-\theta_f/2} \max_{A' \in (\mathbb{Z}/\widetilde{W}\mathbb{Z})^*} S_{|f_1| * \dots * |f_t|}(T; \widetilde{W}; A'). \end{aligned}$$

This completes the proof of Lemma 8.6 as well as the proof of Proposition 5.3.  $\square$

## APPENDIX A. EXPLICIT BOUNDS ON THE CORRELATION OF $\Lambda$ WITH NILSEQUENCES

The aim of this appendix is to provide a proof of Lemma 8.5. This result is due to Green and Tao and we expect that a statement like Lemma 8.5 will eventually appear in [10]. The author is grateful to Ben Green for very helpful discussions.

The proof of Lemma 8.5 rests upon the decomposition of  $\Lambda'$  that already appeared in the proof of the original result, [12, Proposition 10.2]. To be precise, let  $\gamma = \frac{1}{10}2^{-s}$ , where  $s$  is the step of  $G/\Gamma$  and let  $\chi^b + \chi^\sharp = \text{id}_{\mathbb{R}}$  be a smooth decomposition of the identity function  $\text{id}_{\mathbb{R}} : \mathbb{R} \rightarrow \mathbb{R}$  defined via  $\text{id}_{\mathbb{R}}(t) := t$ , such that  $\text{supp}(\chi^\sharp) \subset (-1, 1)$  and  $\text{supp}(\chi^b) \subset \mathbb{R} \setminus [-1/2, 1/2]$ . This decomposition of  $\text{id}_{\mathbb{R}}$  induces the following decomposition of  $\Lambda'$ :

$$\frac{\phi(Wq')}{Wq'} \Lambda'(Wq'n + b') - 1 = \frac{\phi(Wq')}{Wq'} \Lambda^b(Wq'n + b') + \left( \frac{\phi(Wq')}{Wq'} \Lambda^\sharp(Wq'n + b') - 1 \right),$$

where, cf. [12, (12.2)],

$$\Lambda^\sharp(n) = -\log x^\gamma \sum_{d|n} \mu(d) \chi^\sharp\left(\frac{\log d}{\log x^\gamma}\right) \quad (|t| \geq 1 \Rightarrow \chi^\sharp(t) = 0)$$

is a truncated divisor sum, where

$$\Lambda^b(n) = -\log x^\gamma \sum_{d|n} \mu(d) \chi^b\left(\frac{\log d}{\log x^\gamma}\right) \quad (|t| \leq 1/2 \Rightarrow \chi^b(t) = 0)$$

is an average of  $\mu(d)$  running over large divisors of  $n$ . This decomposition in turn splits the correlation from Lemma 8.5 into two correlations that shall be bounded separately.

The correlation estimate of the  $\Lambda^b$  term with nilsequences follows as in [12, §12] from the non-correlation of Möbius with nilsequences and inherits an error term which saves

a factor  $O_A(\log x)^{-A}$  for any given  $A \geq 1$  when compared to the trivial bound. In our setting, we may express the congruence condition modulo  $Wq'$  as a character sum

$$\frac{\phi(Wq')}{Wq'} \Lambda^b(n) \mathbf{1}_{n \equiv b' \pmod{Wq'}} = \mathbb{E}_{\chi \pmod{Wq'}} \frac{\phi(Wq')}{Wq'} \Lambda^b(n) \chi(n) \overline{\chi}(b').$$

Following [12], cf. equation (12.8), the factor  $F(g(n)\Gamma)$  may be reinterpreted as  $F(g'(Wq'n + b')\Gamma)$  for a new polynomial sequence  $g'$ . Reinterpreting the product  $\chi(n)F(g'(n)\Gamma)$  of a character  $\chi$  with the given nilsequence as a nilsequence itself allows us to employ correlation estimate [12, eq. (12.10)] with  $N$  given by  $xq'W \ll x(\log x)^E$  to handle the correlation for  $\Lambda^b$ . Thanks to the saving of an arbitrary power of  $\log x$  in [12, (12.10)], we can compensate the factor of  $Wq'$ , which is bounded above by  $(\log x)^E$ , that we loose when passing to the character sums. In total, we obtain

$$\frac{1}{y} \sum_{n \leq y} \frac{\phi(Wq')}{Wq'} \Lambda^b(Wq'n + b') F(g(n)\Gamma) \ll_{\|F\|_{\text{Lip}, s, G/\Gamma, B}} (\log y)^{-B} \ll_{\|F\|_{\text{Lip}, s, G/\Gamma, B'}} (\log x)^{-B'}.$$

It remains to analyse the contribution of the function  $\lambda^\sharp : \mathbb{N} \rightarrow \mathbb{R}$ , defined via

$$\lambda^\sharp(n) := \frac{\phi(Wq')}{Wq'} \Lambda^\sharp(Wq'n + b') - 1.$$

This contribution satisfies the general bound

$$\left| \frac{1}{y} \sum_{n \leq y} \left( \frac{\phi(Wq')}{Wq'} \Lambda^\sharp(Wq'n + b') - 1 \right) F(g(n)\Gamma) \right| \leq \|\lambda^\sharp\|_{U^{k+1}[y]} \|F(g(\cdot)\Gamma)\|_{U^{k+1}[y]^*}$$

for every  $k \geq 1$ , where the dual uniformity norm is defined via

$$\|F(g(\cdot)\Gamma)\|_{U^{k+1}[N]^*} := \sup \left\{ \left| \frac{1}{N} \sum_{n \leq N} f(n) F(g(n)\Gamma) \right| : \|f\|_{U^k[N]} \leq 1 \right\}.$$

The main task that remains is to obtain control on the above dual uniformity norm for at least one value of  $k$ . In [12], this is achieved through [12, Proposition 11.2], which decomposes a general nilsequence into an averaged nilsequence of bounded dual uniformity norm plus an error term that is small in the  $L^\infty$  norm. The proof of this decomposition uses a compactness argument and, as such, does not provide explicit error terms. Central ideas for a new approach not working with compactness were indirectly provided by work of Eisner and Zorin-Kranich [7] on a different question. Eisner and Zorin-Kranich replace in their work the Lipschitz function in the definition of a nilsequence by a smooth function and the Lipschitz norm by a Sobolev norm. Moreover, they show that certain constructions that play a central role in [13] have counterparts in the Sobolev norm setting. Building on these observations, Green [10] proves that in the Sobolev norm setting the dual  $U^{s+1}$  norm of an  $s$ -step nilsequence is in fact bounded. The statement of the latter result involves the following notion of Sobolev norms.

**Definition A.1** (cf. [10]). Let  $G/\Gamma$  be an  $m$ -dimension nilmanifold together with a Mal'cev basis  $\mathcal{X} = \{X_1, \dots, X_m\}$ . For any  $\psi \in C^\infty(G/\Gamma)$ , set

$$\|\psi\|_{W^m, \mathcal{X}} = \sup_{m' \leq m} \sup_{1 \leq i_1, \dots, i_{m'} \leq m} \|D_{X_{i_1}} \dots D_{X_{i_{m'}}} \psi\|_\infty,$$

where  $D_X \psi(g\Gamma) = \lim_{t \rightarrow 0} \frac{d}{dt} \psi(\exp(tX)g\Gamma)$ .

**Lemma A.2** (Green [10], Theorem 5.3.1). *Let  $G/\Gamma$  be a  $k$ -step nilmanifold together with a filtration  $G_\bullet$  and a  $Q$ -rational Mal'cev basis adapted to it. Let  $g \in \text{poly}(\mathbb{Z}, G_\bullet)$  and suppose  $\tilde{F} \in C^\infty(G/\Gamma)$ . Then*

$$\begin{aligned} \|\tilde{F}(g(\cdot)\Gamma)\|_{U^{k+1}[N]^*} &:= \sup \left\{ \left| \frac{1}{N} \sum_{n \leq N} f(n) \tilde{F}(g(n)\Gamma) \right| : \|f\|_{U^{k+1}[N]} \leq 1 \right\} \\ &\ll Q^{10^k \dim G} \|\tilde{F}\|_{W^{2^k \dim G}, \mathcal{X}}. \end{aligned}$$

In order to apply Lemma A.2 in our situation, an auxiliary result is needed that allows one to pass from the Lipschitz setting to the Sobolev setting, i.e. to write any Lipschitz function on  $G/\Gamma$  as the sum of a smooth function, to which Lemma A.2 can be applied, and a small  $L^\infty$  error. This is the contents of the following lemma which will be proved using a standard smoothing trick; the author thanks Ben Green for pointing out this approach.

**Lemma A.3.** *Suppose that  $F : G/\Gamma \rightarrow \mathbb{C}$  is a Lipschitz function and let  $m$  be a positive integer. Then there is a constant  $c \in (0, 1)$ , only depending on  $G$ , such that for every  $\varepsilon \in (0, c)$  there exists a function  $\psi_m \in C^\infty(G/\Gamma)$  such that*

$$\|F - F * \psi_m\|_\infty \leq \varepsilon(1 + \|F\|_{\text{Lip}}) \quad (\text{A.1})$$

and

$$\|F * \psi_m\|_{W^m, \mathcal{X}} \ll (m/\varepsilon)^{2m} Q^{O(m)}. \quad (\text{A.2})$$

Taking Lemma A.3 on trust for the moment, we first complete the proof of Lemma 8.5 before providing that of Lemma A.3. Recall that the nilmanifold  $G/\Gamma$  from Lemma 8.5 is of step  $s$ . The previous two lemmas allow us to reduce the proof of Lemma 8.5 to a bound on the  $U^{s+1}$ -norm of  $\lambda^\sharp : \mathbb{N} \rightarrow \mathbb{R}$ . More precisely, we have

$$\begin{aligned} &\frac{1}{y} \sum_{n \leq y} \left( \frac{\phi(Wq')}{Wq'} \Lambda^\sharp(Wq'n + b') - 1 \right) F(g(n)\Gamma) \\ &\leq \varepsilon(1 + \|F\|_{\text{Lip}}) + \frac{1}{y} \sum_{n \leq y} \left( \frac{\phi(Wq')}{Wq'} \Lambda^\sharp(Wq'n + b') - 1 \right) (F * \psi_m)(g(n)\Gamma) \\ &\leq \varepsilon(1 + \|F\|_{\text{Lip}}) + \|\lambda^\sharp\|_{U^{s+1}[y]} \|F * \psi_m\|_{U^{s+1}[y]^*}. \end{aligned}$$

Since  $\Lambda^\sharp$  is a truncated divisor sum, one can analyse its  $U^{s+1}$ -norm with the help of [12, Theorem D.3]. We will follow the final paragraph ('The correlation estimate for  $\Lambda^\sharp$ ') of [12, Appendix D] closely.

For each non-empty subset  $\mathcal{B} \subset \{0, 1\}^{s+1}$ , let

$$\Psi_{\mathcal{B}}(n, \mathbf{h}) = \left( Wq'(n + \boldsymbol{\omega} \cdot \mathbf{h}) + b' \right)_{\omega \in \mathcal{B}}, \quad (n, \mathbf{h}) \in \mathbb{Z} \times \mathbb{Z}^{s+1},$$

denote the relevant system of forms. The set of exceptional primes for this system, denoted by  $\mathcal{P}_{\Psi_{\mathcal{B}}}$ , is defined to be the set of all primes  $p$  such that the reduction modulo  $p$  of  $\mathcal{P}_{\Psi_{\mathcal{B}}}$  contains two linearly dependent forms or a form that degenerates to a constant. It is clear that whenever  $x$  is sufficiently large, the set  $\mathcal{P}_{\Psi_{\mathcal{B}}}$  consists of all prime factors of  $W(x)q'$  and, in particular, it contains all primes up to  $w(x)$ . For each prime  $p$ , the local factor  $\beta_p^{(\mathcal{B})}$  corresponding to  $\Psi_{\mathcal{B}}$  is defined to be

$$\beta_p^{(\mathcal{B})} = \frac{1}{p^{2(s+1)}} \sum_{(n, \mathbf{h}) \in (\mathbb{Z}/p\mathbb{Z})^{s+2}} \prod_{\omega \in \mathcal{B}} \frac{p}{\phi(p)} \mathbf{1}_{p \nmid Wq'(n + \boldsymbol{\omega} \cdot \mathbf{h}) + b'}.$$

By [12, Lemma 1.3], we have  $\beta_p^{(\mathcal{B})} = 1 + O_s(1/p^2)$  for all  $p \notin \mathcal{P}_{\Psi_{\mathcal{B}}}$ , and hence

$$\prod_{p \notin \mathcal{P}_{\Psi_{\mathcal{B}}}} \beta_p^{(\mathcal{B})} = 1 + O_s\left(\frac{1}{w(x)}\right) = 1 + O_s\left(\frac{1}{\log \log x}\right),$$

while the product of exceptional local factors satisfies

$$\prod_{p \in \mathcal{P}_{\Psi}} \beta_p^{(\mathcal{B})} = \left( \frac{W(x)q'}{\phi(W(x))q'} \right)^{|\mathcal{B}|}.$$

Let  $K_y$  be a convex body that is contained in the hypercube  $[-y, y]^{s+2}$ , and recall that  $\gamma = \frac{1}{10}2^{-s}$ . Then, [12, Theorem D.3] implies that

$$\begin{aligned} & \frac{1}{y^{s+2}} \sum_{(n, \mathbf{h}) \in K_y} \prod_{\omega \in \mathcal{B}} \Lambda^{\#}(Wq'(n + \boldsymbol{\omega} \cdot \mathbf{h}) + b') \\ &= \frac{\text{vol}(K_y)}{y^{s+2}} \prod_p \beta_p^{(\mathcal{B})} + O_s\left((\log y^{\gamma})^{-1/20} \exp\left(O_s\left(\sum_{p \in P_{\Psi_{\mathcal{B}}}} p^{-1/2}\right)\right)\right). \end{aligned}$$

Since  $Wq' \leq (\log x)^E$ , we have  $|P_{\Psi_{\mathcal{B}}}| \ll \frac{w(x)}{\log x} + \frac{E \log \log x}{\log w(x)}$ . Recall that  $w(x) \leq \log \log x$  and that  $\log y \in [(\log x)^{\alpha}, \log x]$ . Thus,

$$\begin{aligned} (\log y^{\gamma})^{-1/20} \exp\left(O_s\left(\sum_{p \in P_{\Psi_{\mathcal{B}}}} p^{-1/2}\right)\right) &\ll (\gamma(\log x)^{\alpha})^{-1/20} \exp\left(O_s(|P_{\Psi_{\mathcal{B}}}|)\right) \\ &\ll_s (\log x)^{-\alpha/20} (\log x)^{O(E)/\log w(x)}, \end{aligned}$$

which is  $o(1)$  as  $x \rightarrow \infty$ .

Choosing  $K_y = \{(n, \mathbf{h}) : 0 < n + \boldsymbol{\omega} \cdot \mathbf{h} \leq y \text{ for all } \boldsymbol{\omega} \in \{0, 1\}^{s+1}\}$ , we obtain

$$\begin{aligned} \|\lambda^\sharp\|_{U^{s+1}[y]}^{2^{s+1}} &= \frac{\text{vol}(K_y)}{y^{s+2}} \sum_{\mathcal{B} \subseteq \{0,1\}^{s+1}} (-1)^{|\mathcal{B}|} \prod_{p \notin P_{\Psi_{\mathcal{B}}}} \beta_p^{(\mathcal{B})} + O_s\left((\log x)^{-\alpha/20 + \frac{O_s(E)}{\log w(x)}}\right) \\ &\ll_s \frac{\text{vol}(K_y)}{y^{s+2}} \frac{1}{\log \log x} + (\log x)^{-\alpha/20 + \frac{O(E)}{\log w(x)}} \\ &\ll_{s,\alpha,E} \frac{1}{\log \log x} \end{aligned}$$

Applying Lemma A.3 with  $m = 2^s \dim G$  and  $\varepsilon = (\log \log x)^{-1/4m}$ , it follows from Lemma A.2 that for  $\exp((\log x)^\alpha) \leq y \leq x$

$$\begin{aligned} &\frac{1}{y} \sum_{n \leq y} \left( \frac{\phi(W)}{W} \Lambda'(Wq'n + b') - 1 \right) F(g(n)\Gamma) \\ &\leq \frac{1 + \|F\|_{\text{Lip}}}{(\log \log x)^{1/(2^{s+2} \dim G)}} + \frac{1}{y} \sum_{n \leq y} \left( \frac{\phi(W)}{W} \Lambda'(Wq'n + b') - 1 \right) (F * \psi_m)(g(n)\Gamma) \\ &\ll_{s,\alpha,E} \frac{1 + \|F\|_{\text{Lip}}}{(\log \log x)^{1/(2^{s+2} \dim G)}} + \left\| \frac{\phi(W)}{W} \Lambda^\sharp(Wq'n + b') - 1 \right\|_{U^{s+1}[y]} \left\| F * \psi_m \right\|_{U^{s+1}[y]^*} \\ &\ll_{s,\alpha,E} \frac{1 + \|F\|_{\text{Lip}}}{(\log \log x)^{1/(2^{s+2} \dim G)}} + \frac{(\log \log x)^{1/2} Q^{10^s \dim G}}{\log \log x}, \end{aligned}$$

which reduces the proof of Lemma 8.5 to that of Lemma A.3.

*Proof of Lemma A.3.* Let  $d_{\mathcal{X}}$  denote the metric on  $G/\Gamma$  that was introduced in [13, Definition 2.2] and define for every  $\varepsilon' > 0$  the following  $\varepsilon'$ -neighbourhood

$$\mathcal{B}_{\varepsilon'} = \{x \in G/\Gamma : d_{\mathcal{X}}(x, \text{id}_G \Gamma) < \varepsilon'\}.$$

Let  $\varepsilon \in (0, 1)$ . Since  $F$  is Lipschitz, we have  $|F(x) - F(y)| \leq \varepsilon(1 + \|F\|_{\text{Lip}})$  whenever both  $x$  and  $y$  belong to the neighbourhood  $\mathcal{B}_\varepsilon$  of  $\text{id}_G \Gamma$ . To ensure that (A.1) holds, it thus suffices to ensure that  $\psi_m$  is non-negative, supported in  $\mathcal{B}_\varepsilon$  and that  $\int_{G/\Gamma} \psi_m = 1$ . Indeed, these assumptions imply that

$$\begin{aligned} \left| F(x) - \int_{G/\Gamma} F(y) \psi_m(x - y) dy \right| &= \left| \int_{G/\Gamma} (F(y) - F(x)) \psi_m(x - y) dy \right| \\ &\leq \varepsilon(1 + \|F\|_{\text{Lip}}) \int_{G/\Gamma} \psi_m(x - y) dy = \varepsilon(1 + \|F\|_{\text{Lip}}). \end{aligned}$$

The function  $\psi_m$  will be constructed as the  $m$ -fold convolution of a smooth bump-function. For this purpose, observe that

$$m\mathcal{B}_{\varepsilon/m} \subseteq \mathcal{B}_\varepsilon.$$

If  $g = \exp(s_1 X_1) \dots \exp(s_{\dim G} X_{\dim G})$ , then the (unique) coordinates

$$\psi(g) := (s_1, \dots, s_{\dim G})$$

are called Mal'cev coordinates, while the unique coordinates

$$\psi_{\exp}(g) := (t_1, \dots, t_{\dim G})$$

for which  $g = \exp(t_1 X_1 + \dots + t_{\dim G} X_{\dim G})$  are called exponential coordinates. Proceeding as in the proof of [13, Lemma A.14], one can identify  $G/\Gamma$  with the fundamental domain  $\{g \in G : \psi(g) \in [-\frac{1}{2}, \frac{1}{2})\} \subset G$ . Furthermore, [13, Lemma A.2] shows that the change of coordinates between exponential and Mal'cev coordinates, i.e.  $\psi \circ \psi_{\exp}^{-1}$  or  $\psi_{\exp} \circ \psi^{-1}$ , is in either direction a polynomial mapping with  $Q^{O(1)}$ -rational coefficients. Thus,  $\mathcal{B}_\varepsilon$  lies within the fundamental domain provided  $\varepsilon < c_0$  for some sufficiently small constant  $c_0$ . This embedding of  $\mathcal{B}_\varepsilon$  in  $G$  allows us to define  $\log$  on  $\mathcal{B}_\varepsilon$ . Let us equip  $\mathfrak{g}$  with the maximum norm associated to  $\mathcal{X}$ , that is  $\|X\| := \max_i |t_i|$  for  $X = \sum_i t_i X_i$ . Then the definition of  $d_{\mathcal{X}}$  and [13, Lemma A.2] imply that

$$\{X \in \mathfrak{g} : \|X\| < \delta\} \subseteq \log \mathcal{B}_{\varepsilon/m}$$

for some  $\delta$  of the form  $\delta = \frac{\varepsilon}{m} Q^{-O(1)}$ . Following the above preparation, we now choose a non-negative smooth function  $\chi_1 : \mathbb{R}^{\dim G} \rightarrow \mathbb{R}_{\geq 0}$  with support in  $\{\mathbf{t} \in \mathbb{R}^{\dim G} : \|\mathbf{t}\|_\infty < 1\}$  that satisfies  $\int_{\mathbb{R}^{\dim G}} \chi_1(\mathbf{t}) d\mathbf{t} = 1$ . Then, by setting  $\chi(\mathbf{t}) = \delta \cdot \chi_1(\delta \mathbf{t})$ , we obtain a function  $\chi : \mathbb{R}^{\dim G} \rightarrow \mathbb{R}_{\geq 0}$  that is supported on  $\{\mathbf{t} \in \mathbb{R}^{\dim G} : \|\mathbf{t}\|_\infty < \delta\}$ , satisfies  $\int_{\mathbb{R}^{\dim G}} \chi(\mathbf{t}) d\mathbf{t} = 1$  and has furthermore the property that

$$\left\| \frac{\partial}{\partial t_i} \chi(t_1, \dots, t_{\dim G}) \right\|_\infty \ll (m/\varepsilon)^2 Q^{O(1)} \quad (\text{A.3})$$

for  $1 \leq i \leq \dim G$ . We may identify  $\chi$  with a function defined on the vector space  $\mathfrak{g}$  equipped with the basis  $\{X_1, \dots, X_{\dim G}\}$ , by setting  $\chi(t_1 X_1 + \dots + t_{\dim G} X_{\dim G}) = \chi(t_1, \dots, t_{\dim G})$ .

To obtain a smooth bump-function on  $G/\Gamma$ , we consider the composition  $\chi \circ \log : G/\Gamma \rightarrow \mathbb{R}$ , which is supported in  $\mathcal{B}_{\varepsilon/m}$ . Since the differential  $d \log_{\text{id}_G} : \mathfrak{g} \rightarrow \mathfrak{g}$  is the identity, there are positive constants  $C_0, C_1$  and  $c_1$ , such that

$$C_0 \leq \int_{G/\Gamma} \chi \circ \log \leq C_1,$$

provided  $\varepsilon < c_1$ . Hence there is a constant  $C$  such that  $\int_{G/\Gamma} \psi = 1$  for  $\psi = C\chi \circ \log$ .

With this function  $\psi$  at hand, let  $\psi_m = \psi^{*m}$  be the  $m$ -th convolution power of  $\psi$ . It is clear that for every  $0 < k \leq m$ , the function  $\psi^{*k}$  is supported in  $\mathcal{B}_\varepsilon$  and that  $\int_{G/\Gamma} \psi^{*k} = 1$ . Setting  $\psi^{*0} = \delta_0$ , where  $\delta_0$  denotes the Kronecker  $\delta$ -function with weight 1 at 0, we furthermore have

$$D_{X_{i_1}} \dots D_{X_{i_k}} (F * \psi_m) = F * D_{X_{i_1}} \psi * \dots * D_{X_{i_k}} \psi * \psi^{*(m-k)}$$

and, hence,

$$\|D_{X_{i_1}} \dots D_{X_{i_k}} (F * \psi_m)\|_\infty \leq \|F\|_\infty \cdot \|D_{X_{i_1}} (C\chi \circ \log)\|_\infty \dots \|D_{X_{i_k}} (C\chi \circ \log)\|_\infty$$

for any  $k \leq m$ . Our final task is to bound  $\|D_{X_j}(C\chi \circ \log)\|_\infty$  for every  $j \leq \dim G$ . Writing  $[\cdot]_i : \mathfrak{g} \rightarrow \mathbb{R}$  for the  $i$ -th co-ordinate map with respect to the basis  $\mathcal{X}$ , we have

$$D_{X_j}(\chi \circ \log)(g) = \sum_{i=1}^{\dim G} \frac{\partial \chi}{\partial X_i}(\log g) \cdot \lim_{t \rightarrow 0} \left[ \log(\exp(tX_j)g) \right]_i. \quad (\text{A.4})$$

Since the differential  $d\log_{\text{id}_G} : \mathfrak{g} \rightarrow \mathfrak{g}$  is the identity, there are constants  $C_1 > 0$  and  $c_1 > 0$ , such that for every  $g \in \mathcal{B}_{c_1}$  and for  $1 \leq i \leq m$ , the derivative

$$\left| \lim_{t \rightarrow 0} \left[ \log(\exp(tX_j)g) \right]_i \right|$$

is bounded by  $C_1$ . Choosing  $c < \min(c_0, c_1, c_2)$ , the bound (A.2) now follows from (A.4) and the bounds given in (A.3).  $\square$

## REFERENCES

- [1] G. Bachman, On exponential sums with multiplicative coefficients, II. *Acta Arith.* **106** (2003), 41–57.
- [2] J. Bourgain, P. Sarnak and T. Ziegler, Disjointness of Moebius from Horocycle Flows. In: *From Fourier Analysis and Number Theory to Radon Transforms and Geometry*, Dev. Math., vol. 28, Springer, New York, 2013, 67–83.
- [3] T.D. Browning and L. Matthiesen, Norm forms for arbitrary number fields as products of linear polynomials, arXiv:1307.7641, 2013.
- [4] H. Daboussi and H. Delange, Quelques propriétés des fonctions multiplicatives de module au plus égal à 1, *C.R. Acad. Sci. Paris, Sér. A*, **278** (1974), 657–660.
- [5] H. Daboussi and H. Delange, On multiplicative arithmetical functions whose modulus does not exceed one. *J. London Math. Soc.* (2), **26** (1982), 245–264.
- [6] Y. Dupain, R.R. Hall and G. Tenenbaum, Sur l’équidistribution modulo 1 de certaines fonctions de diviseurs. *J. London Math. Soc.* (2), **26** (1982), 397–411.
- [7] T. Eisner and P. Zorin-Kranich, Uniformity in the Wiener–Wintner theorem for nilsequences *Discrete Contin. Dyn. Syst.* **33** (2013) 3497–3516.
- [8] P. Erdős, On the sum  $\sum_{k=1}^x d(f(k))$ , *J. London Math. Soc.* **27** (1952) 7–15.
- [9] N. Frantzikinakis and B. Host, Higher order Fourier analysis of multiplicative functions and applications, arXiv:1403.0945, 2014.
- [10] B.J. Green, *Hadamard Lectures on Nilsequences*. Book in preparation.
- [11] B.J. Green and T.C. Tao, The primes contain arbitrarily long arithmetic progressions. *Annals of Math.* **167** (2008), 481–547.
- [12] B.J. Green and T.C. Tao, Linear equations in primes. *Annals of Math.* **171** (2010), 1753–1850.
- [13] B.J. Green and T.C. Tao, The quantitative behaviour of polynomial orbits on nilmanifolds. *Annals of Math.* **175** (2012), 465–540.
- [14] B.J. Green and T.C. Tao, The Möbius function is strongly orthogonal to nilsequences. *Annals of Math.* **175** (2012), 541–566.
- [15] B.J. Green, T.C. Tao and T. Ziegler, An inverse theorem for the Gowers  $U_{s+1}[N]$ -norm. *Annals of Math.* **176** (2012), 1231–1372.
- [16] A.J. Harper, A different proof of a finite version of Vinogradov’s bilinear sums inequality, unpublished notes, 2011. Available at <https://www.dpmms.cam.ac.uk/~ajh228/FiniteBilinearNotes.pdf>
- [17] L.K. Hua, *Additive Theory of Prime Numbers*, AMS, 1965.
- [18] A.J. Irving, The divisor function in arithmetic progressions to smooth moduli, *Int. Math. Res. Not. IMRN*, to appear.
- [19] I. Kátai, A remark on a theorem of H. Daboussi, *Acta Mathematica Hungarica*, **47** (1986) 223–225.
- [20] L. Matthiesen, Correlations of the divisor function. *Proc. London Math. Soc.* **104** (2012), 827–858.



- [21] L. Matthiesen, Linear correlations amongst numbers represented by positive definite binary quadratic forms. *Acta Arith.* **154** (2012), 235–306.
- [22] H.L. Montgomery and R.C. Vaughan, Exponential sums with multiplicative coefficients. *Invent. Math.* **43** (1977), 69–82.
- [23] M. Nair and G. Tenenbaum, Short sums of certain arithmetic functions. *Acta Math.* **180** (1998), 119–144.
- [24] K. Prachar, Über Zahlen der Form  $a^2 + b^2$  in einer arithmetischen Progression. *Math. Nachr.* **10** (1953), 51–54.
- [25] P. Shiu, A Brun–Titchmarsh theorem for multiplicative functions, *J. reine angew. Math.* **313** (1980), 161–170.
- [26] T. Tao, The Katai–Bourgain–Sarnak–Ziegler orthogonality criterion, blog post, 2011. Available at <http://terrytao.wordpress.com/2011/11/21/the-bourgain-sarnak-ziegler-orthogonality-criterion/>
- [27] E. Wirsing, Das asymptotische Verhalten von Summen über multiplicative Funktionen. II, *Acta Math. Acad. Sci. Hungar.* **18** (1967), 411–467.

INSTITUT DE MATHÉMATIQUES DE JUSSIEU — PARIS RIVE GAUCHE, UMR 7586, BÂTIMENT SOPHIE GERMAIN, CASE 7012, 75205 PARIS CEDEX 13, FRANCE

*E-mail address:* lilian.matthiesen@imj-prg.fr